STATISTICAL LEARNING WITH REGULARIZATIONS: THEORY AND APPLICATIONS

A Dissertation
Presented to
The Academic Faculty

By

Shanshan Cao

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology

August 2019

STATISTICAL LEARNING WITH REGULARIZATIONS: THEORY AND APPLICATIONS

Approved by:

Dr. Xiaoming Huo, Advisor
H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology

Dr. Nicoleta Serban, Co-Advisor
H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology

Dr. Joel Sokol
H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology

Dr. Branislav Vidakovic
H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology

Dr. Vladimir Koltchinskii
School of Mathematics
Georgia Institute of Technology

Date Approved: May 3, 2019

Dedicated to my parents, Mei and Shuen, my parents-in-law, Hongwen and Qingyu,

my husband, Ruizhi, and my sister, Danyang,

for their boundless love and support

Shanshan Cao

ACKNOWLEDGEMENTS

Like a miracle, 5 years ago, when I first arrived at Atlanta, I was so nervous about the survival as a Ph.D. student; now I am doing exciting cutting-edge research in statistical learning and applications. They are my advisors, Prof. Xiaoming Huo and Prof. Nicoleta Serban, who make that happen. I would like to thank Dr. Xiaoming Huo for his guidance and endless patience over the past four years as my dissertation advisor and mentor. I would like to thank Dr. Nicoleta Serban for her support and guidance during my PhD study, especially the first three years on my research, my life and my career goals. Along the journey of my research pursuit, their visionary thought and insightful guidance help me walk through hard times. It is my best fortune to have been working with them and they have been and will always be role models through my life.

Thanks to the professors who spend their precious time to serving on my dissertation committee: Prof. Vladimir Koltchinskii, Prof. Joel Sokol and Prof. Branislav Vidakovic. I have to specially thank Prof. Vladimir Koltchinskii for teaching me two graduate math courses on hypothesis testing and statistical estimation, which helped me build a solid foundation and benefit me a lot in my research and future career and I enjoyed the math flavor courses a lot; Prof. Joel Sokol for his suggestions and guidance during my serving as his Teaching Assistant for creating an online course in the Analytics program, where I learned a lot of skills on being a good instructor, especially for online courses; and Prof. Branislav Vidakovic for his valuable suggestions and comments during the preparation of my dissertation.

I also want to express my gratitude to Prof. Monica Gentili, Prof. Yajun Mei, Prof. C. F. Jeff Wu, Prof. Robert Foley, Prof. Paul Griffin, Prof. Yao Xie, Prof Jong-Shi, Pang, Prof. Jianjun Shi, Prof. George V. Moustakides, who provide invaluable suggestions on my research directions and career, and Prof. Sigrun Andradottir, Prof. Hayriye Ayhan, Prof. Alan Erera, Prof. Roshan Joseph, Prof. Enlu Zhou, Prof. Arkadi Nemirovski, Prof.

Santosh Vempala, who gave great lectures during my PhD study.

During the pursuit of my Ph.D. at Georgia Tech these years, I owe my thanks to many people. Thanks to Qingqing Liu and her husband Rundong Du for their selfless help, especially they provided us place to stay in our first day to the US. Thanks to Yuanshuo Zhao and Jing Qin for being my most trustable colleagues and friends during my PhD study, and more importantly, best man in my wedding. Thanks to Junqing Qian for being my best friend since undergraduate, who almost made it to be my bridesmaid in my wedding, and visited me several times in Atlanta during my stay here. I would also extend my thanks to other friends and colleagues in Georgia Tech, who make my PhD less stressful and colorful, and thank May Li, Kathy Huggins, Amanda Ford, etc., for their great administrative support.

I am very fortunate to work as an intern in IBM in the summer 2016 and meet excellent researchers there. Thanks to Kun Hu for her great mentorship.

Finally, I am greatly indebted to my parents for their boundless love and support and my little sister for her company. I am thankful to my parents-in-law, for their support and care. I want to express my deepest gratitude to my dear husband, Ruizhi, who has always been together with me for more than 9 years and will always be with me in my future journey. Besides his stressful thesis work, he provided me suggestions, support, patience and love. Without their support and sacrifice, I wouldn't have these amazing accomplishments. I will always be with them for the unknown but anticipated journey in the rest of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This thesis contributes to the area of statistical learning with regularization, which has been popular for sparse estimation and function estimation in many areas such as signal/image processing, statistics, bioinformatics and machine learning. Our study helps (i) unify the framework of high-dimensional sparse estimation with non-convex penalty; (ii) prove the asymptotical optimality of high-order Laplacian regularization in function estimation; (iii) improve the performance of the composite fuselage assembly process by using sparsity penalized ℓ_∞ based linear model; (iv) identify the census tracts where children have limited access to preventive dental care.

In this thesis, we have four main works. In Chapter 1, under the linear regression framework, we study the variable selection problem when the underlying model is assumed to have a small number of nonzero coefficients (i.e., the underlying linear model is sparse). Non-convex penalties in specific forms are well-studied in the literature for sparse estimation. A recent work [1] has pointed out that nearly all existing non-convex penalties can be represented as difference-of-convex (DC) functions, which can be expressed as the difference of two convex functions, while itself may not be convex. There is a large existing literature on the optimization problems when their objectives and/or constraints involve DC functions. Efficient numerical solutions have been proposed. Under the DC framework, directional-stationary (d-stationary) solutions are considered, and they are usually not unique. In this chapter, we show that under some mild conditions, a certain subset of d-stationary solutions in an optimization problem (with a DC objective) has some ideal statistical properties: namely, asymptotic estimation consistency, asymptotic model selection consistency, asymptotic efficiency. The aforementioned properties are the ones that have been proven by many researchers for a range of proposed non-convex penalties in the sparse estimation. Our assumptions are either weaker than or comparable with those conditions that have been adopted in other existing works. This work shows that DC is a nice

framework to offer a unified approach to these existing work where non-convex penalty is involved. Our work bridges the communities of optimization and statistics.

In Chapter 2, we propose a function estimation method using the high-order Laplacian regularization. Graph Laplacian based regularization has been widely used in learning problems to take advantage of the information on the geometry towards the marginal distribution. In this chapter, we consider the high-order Laplacian regularization, whose empirical (i.e., sample) version takes the form of $\mathbf{f}^T\mathbf{L}^m\mathbf{f}$ with $\mathbf{L}$ being the graph Laplacian matrix of the sample data, in the context of supervised learning, and provide the theoretical foundations in the non-parametric setting. We call the resulting estimator a *Graph Laplacian Smoother (GLS)*. The high-order Laplacian regularization technique, which is proved to converge to the Sobolev semi-norm regularization, has been successfully used in the literature of semi-supervised learning and supervised learning problems without theoretical guarantees. In this work, it is shown that nearly all good asymptotic properties of the existing state-of-the-art approaches are inherited by the Laplacian-based smoother. Specifically, we prove that as the sample size goes to infinity, the expected mean squared errors (MSE) is of order $O(n^{-\frac{2m}{2m+d}})$, which is the *optimal convergence rate* in a setting of nonparametric estimation [2], where m is the order of the Sobolev semi-norm used in the regularization, and d is the intrinsic dimension of the domain. Besides, we propose a *generalized cross validation* (GCV) approach to choose the penalty parameter λ, and we establish its *asymptotical optimality* guarantee.

In Chapter 3, we study the fuselage assembly problem using sparse learning theories. Natural dimensional variabilities of incoming fuselages affect the assembly speed and quality of fuselage joins in composite fuselage assembly process. Thus, shape control is critical to ensure the quality of composite fuselage assembly. In current practice, the structures are adjusted to the design shape in terms of ℓ_2 loss for further assembly without considering the initial dimensional gap between two structures. Such practice has two limitations: (1) the design shape may not be the optimal shape in terms of a pair of incoming fuselages

with different incoming dimensions; (2) the maximum gap is the key concern during the fuselage assembly process, so the ℓ_∞ loss of gap after control needs to be considered. This paper proposes an optimal shape control methodology via ℓ_∞ loss for composite fuselage assembly process by considering the initial dimensional gap between the incoming pair of fuselages. Due to the limitation on the number of available actuators in practice, we face an important problem of finding the best locations for the actuators among many potential locations, which makes our problem a sparse estimation problem. We are the first to solve optimal shape control in fuselage assembly process using ℓ_∞ model under the framework of high-dimensional sparse estimation, where we use the ℓ_1 penalty to control the sparsity of the resulting estimator. From statistical point of view, this can be formulated as the ℓ_∞ loss based linear regression, and under some standard assumptions, such as restricted eigenvalue (RE) conditions, and the light tailed noise, the non-asymptotic estimation error of the ℓ_1 regularized ℓ_∞ linear model in the order of $O(\sigma\sqrt{\frac{S\log p}{n}})$, is derived, which meets the upper-bound in the existing literature. Compared to the current practice, the case study shows that our proposed method significantly reduces the maximum gap between two fuselages after shape adjustments by using comparable forces.

In Chapter 4, we compared access to preventive dental care for low-income children eligible for public dental insurance to children with private dental insurance and/or high family income (>400% of the federal poverty level) in Georgia and the impact of policies towards increasing access to dental care for low-income children. Specifically, we used multiple sources of data (e.g., US Census, Georgia Board of Dentistry) to estimate measures of preventive care access in 2015 for children, aged 0 to 18 years. Measures included met need, scarcity of dentists, and one-way travel distance to a dentist at the census tract level. We used an optimization model to estimate access, quantify disparities and evaluate policies. We find that about 1.5 million children were eligible for public insurance, and 600,000 had private insurance and/or high family income. Across census tracts, average met need was 59% for low-income children and 96% for the high-income children; for

rural census tracts, these values were 33% and 84%, respectively. The average travel distance for all census tracts was 3.71 miles for high-income/insured children and 17.16 miles for low-income children; for rural census tracts, these values were 11.55 and 32.91 miles, respectively. Met need significantly increased and travel distance decreased for modest increases in provider acceptance of Medicaid eligible children. In order to achieve 100% met need, 80% provider participation rate would be required. We conclude that across census tracts, high-income children had notably higher access than low-income children. Identifying these tracts could result in more efficient allocation of public health dental resources.

CHAPTER 1

A UNIFYING FRAMEWORK OF HIGH-DIMENSIONAL SPARSE ESTIMATION WITH DIFFERENCE-OF-CONVEX (DC) REGULARIZATIONS

1.1 Introduction

Sparse estimation under a linear regression model is a fundamental and classical problem in statistics. It continues to be highly active in the high-dimensional regime when the underlying parameter is believed to be sparse. Properties on the resulting estimators have been extensively studied with different penalties of the sparsity in [3, 4, 5, 6, 7, 8, 9, 10, 11], etc. However, most existing works focus on the properties on a specific solution to the possibly nonconvex objective function, which is used to derive a sparse estimation of the unknown parameter. The stationary solutions of other kind might also be of interest and possess satisfying properties, such as the desired asymptotic estimation consistency, asymptotic model selection consistency, asymptotic efficiency. A unified framework for the penalized high-dimensional sparse estimation and the relation to a subfield of optimization problems, namely, the difference-of-convex (DC) programming are missing in the literature. We establish such a connection in this chapter.

1.1.1 Sparsity induced penalties

We first present the formulation of high-dimensional sparse estimation in linear regression setting using sparsity induced penalties. Consider observations (y_1, x_1), (y_2, x_2), ..., (y_n, x_n), where we have the response $y_i \in \mathbb{R}$ and the predictor $x_i \in \mathbb{R}^p$ satisfy

$$y_i = \beta^{*T} x_i + \epsilon_i.$$

Here, β^{*T} is the transpose of the vector $\beta^* \in \mathbb{R}^p$, which is the true however unknown underlying parameter to be estimated. We further assume that noises ϵ_i's are independently distributed, with 0 mean and equal variance σ^2 (which can be a sub-Gaussian distribution with variance parameter σ^2), and are independent of x_i's. The above model is commonly written in the following matrix form:

$$y = X\beta^* + \epsilon, \tag{1.1.1}$$

where the vector $y = (y_1, \cdots, y_n)^T \in \mathbb{R}^n$ is the response vector, $X \in \mathbb{R}^{n \times p}$ is the model matrix with rows being individual predictors, $x_1^T, \cdots, x_n^T$, and the random vector ϵ contains the noises.

In the high-dimensional regime where the number of the parameters, denoted by p, exceeds the sample size, denoted by n, one of the most important methods (according to many works such as [12, 13, 14]) is to estimate the parameter by using the LASSO [15] approach. It is interesting to note that a mathematically equivalent approach was proposed in [16] around the same time in the computational and applied mathematics literature. LASSO is defined through solving the following convex optimization problem:

$$\hat{\beta}^{lasso}(y, X; \lambda) = \arg\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \tag{1.1.2}$$

The first term in the above objective is the goodness-of-fit measure (a.k.a., the residual sum of squares) in the linear regression model (1.1.1). The second term in the objective is a penalty function, which is the sum of absolute values: $\sum_{i=1}^k \lambda|\beta_i|$. We can further write the penalty in a more general form $\sum_{i=1}^k P_\lambda(\beta_i)$, where the univariate function $P_\lambda(x)$ takes the form $P_\lambda(x) = \lambda|x|$ in the LASSO approach. Many existing works, including [3] and [4] and others, have proved that with high probability (i.e., probability goes to 1 as sample size goes to infinity), under some conditions on the design matrix and the choices for λ, the LASSO will be able to find the right signed support for the unknown parameter β^*.

The cases that have been studied include (1) when the matrix can be fixed or random, (2) the dimension of the unknown parameter is fixed or goes to infinity as the sample size increases, and (3) other interesting situations.

Despite the success of obtaining sparse estimators by using LASSO, it is also well known that the resulting estimator is biased. This can be readily seen by considering the special scenario, where the design matrix X is orthonormal, consequently the L_1 penalty leads to a soft-thresholding solution, which is biased from the true parameter β^*. De-biasing procedure has been studied in [17, 18, 19, 20]. In the present chapter, we decide to focus on the regularization (i.e., adding a penalty function) approach, partially because the de-biasing approach may require solving multiple optimization problems, therefore could be computationally disadvantageous. At the same time, we may explore the other algorithmic-design approaches in the future.

An effective extension of the LASSO estimator is to replace the penalty function $P_\lambda(x)$ in (1.1.2) into some folded concave functions, which are non-convex. Some representitive works include SCAD [5, 6], MCP [9], adaptive LASSO [7], capped-l_1 [11], together with others. In general, this leads to an NP-hard problem; therefore no polynomial-time algorithm is known in finding the global optimal solution. Specifically, SCAD is proposed in [5, 6], in order to debias the estimation when the parameter is numerically relatively large, which gives a constant penalty as the parameter is large enough. Adaptive LASSO is studied in [7, 8], which is motivated by the fact that in the orthogonal design, the bias of the parameter estimation is approximately λ in LASSO. The authors suggest to give different sizes of penalties to different parameters, so that the variables with large coefficients have smaller weights in the ℓ_1 penalty (depending on some consistent estimator $\hat{\beta}$ of β^*). Then they can reduce the estimation bias of the lasso, while retaining its sparsity property. MCP is proposed in [9], which also gives a constant penalty as the parameter is large enough. Capped-l_1 in [10, 11] gives a penalty of truncated l_1 penalty to ensure a constant penalty when the estimation is large. Consistency results, including measuring the squared dis-

tance of the estimation, prediction, signed support recovery, for the previously mentioned formulations can be found in the original works.

1.1.2 Difference-of-convex (DC) unified penalties

Recently, it is pointed out by [1] that all the previously listed penalties can be written in a unified DC form. Especially, the first term is the ℓ_1 penalty $\|\beta\|_1$. This leads to a DC formulation; i.e., solving the penalized least square problem with a generalized DC penalty function. In [1], they prove under some strong assumptions (strong convexity of the loss functions) that the d-stationary solutions found by a standard DC algorithm (i.e., DCA) is in fact the global minimal. This result might not be surprising because under their assumptions, the objective function (the DC-penalized loss functions) in fact can be strongly convex, which makes the solution unique. They also prove that the ℓ_0 norm of the d-stationary solution has an upper-bound, which doesn't shed lights on the support recovery property. Our work is inspired by [1], compared to [1], we relax the assumptions on strong convexity for the loss function and prove the existence of a class of d-stationary solutions, which have the oracle properties in the linear regression scenario. Our result indicates that the assumption on the strong convexity of the loss function within the domain is not necessary. In addition, the aforementioned work has an applied mathematics focus – their ℓ_0 norm result does not imply statistical properties of the d-stationary solutions. In statistical literature, the distance between the d-stationary solution and the assumed ground truth is considered. Our result will be more formulated towards the statistical properties of the d-stationary solutions: namely model estimation consistency, asymptotic convergence rate in estimation, and model selection consistency. Despite the difference, it is interesting to point out that both work will require the restricted convexity assumption, which is assumed in nearly all related work. Besides, we also generalize the results to DC penalized general loss functions.

The use of DC functions offers a general framework on non-convex regularization.

Some special cases are discussed in [21] and [22], although they don't explicitly mention the DC functions in their work. The first work [21] assumes that the penalty function $p_\lambda(\beta)$ is separable in parameter β and each of the univariate penalty can be written as the difference of a convex function $p_{\lambda,\mu}(t)$ and a quadratic function $\frac{\mu}{2}t^2$, where μ is a known positive constant. Therefore one has $p_\lambda(t) = p_{\lambda,\mu}(t) - \frac{\mu}{2}t^2$. They restrict the feasible region to a bounded region containing the ground truth β^*. Under certain regularity conditions on the penalty, including differentiability, and restrictive strong convexity of the loss function, they give the optimal upper-bound for the estimation error as well as for prediction error. Their assumption includes the popular studied penalties like SCAD and MCP. On the other hand, They don't have results on the support recovery and they purposely eliminated possible stationary solutions outside the bounded feasible region they defined. The second work [22] mainly assumes the restricted strong convexity of the penalties and the loss functions. They mainly discuss the elliptical design regression, least square loss, and logistic loss with SCAD, MCP penalties which can be written as the summation of the ℓ_1 penalty $\|\beta\|_1$ and a concave function $q_\lambda(\beta)$ with proper bound in the concavity. They argue that the local quadratic approximation algorithm they provided converges to a unique local minimum which enjoys the oracle properties as if you've already know the support for the true parameter. They prove the estimation error upper-bound. And they are able to prove the support recovery results for linear regression model with least square loss function. Both are under the assumption that the concavity of the function $q_\lambda(\beta)$ is bounded.

Both works [21, 22] assume the decreasing first order derivative of the penalty function on the nonnegative real line, which is necessary for the unbiasedness for estimation of larger β. They both restrict the penalties such that the objective function is strongly convex within some region where the local optimal solutions as well as the true unknown parameters are in the given convex set.

While in the current work, we solve the unconstrained problem and prove the asymptotic convergence results of the estimation for a class of local d-stationary solutions without

using the assumption of the bounded concavity of the $q_\lambda(\cdot)$ function and constant penalty when the parameter is large enough, which allows us to include other penalties such as transformed ℓ_1 [23, 24] and logarithmic [25], into our analysis. Equipped with the bounded convexity assumption, we further prove that the support recovery consistency for the class of d-stationary solutions we find near the ground truth.

From the computation perspective, there is a rich literature on solving the penalized (also known as regularized) problem numerically. For example, Local Linear Approximation (LLA) in [26] prove that one-step estimator from LLA performs well in SCAD with penalized likelihood estimation. They also prove the asymptotic normality under some regularity conditions of the Fisher Information matrix. In this chapter, we would like to explore the relationship between LLA and the popular DC Algorithm (DCA) which is often used in DC programming. It turns out that all the above mentioned algorithms are special cases of DCA.

This chapter builds a bridge between optimization, where people focused on solving the optimization problem efficiently, and statistics, where people mainly focused on the inference (finding the estimation). The link here would be the DC programming and DCA. The DC programming enables us to generalize the classical penalized likelihood function to the DC penalized likelihood function, while DCA provides us efficient algorithms to solve the corresponding numerical problems. Borrowing strength from existing literatures enables us to solve the optimization problem efficiently with convergence guarantees. We unify the existing algorithms in the literature for finding the local minima of non-convex optimization problems under the DCA framework.

1.1.3 Notations

For a real number $q \in [1, +\infty)$, the ℓ_q norm of a vector $\beta = (\beta_1, \beta_2, \cdots, \beta_p) \in \mathbb{R}^p$ is defined as $\|\beta\|_q = (\sum_{i=1}^p \beta_i^q)^{1/q}$. Specially, the ℓ_∞ norm is defined as $\|\beta\|_\infty = \max_{i=1}^p\{|\beta_i|\}$. The ℓ_0 norm is defined as $\|\beta\|_0 = \text{card}\{\text{supp}(\beta)\}$, where we have $\text{supp}(\beta) = \{i : \beta_i \neq 0\}$

and card$\{\cdot\}$ is the cardinality of the set. We denote the cardinality of a set S by $|S|$ and its complement by S^c. For $\beta = (\beta_1, \beta_2, \cdots, \beta_p) \in \mathbb{R}^p$, we let β_S denote the sub-vector (of β) whose elements correspond to the set S; we let X_S denote the sub-matrix (of X) whose columns indices are correspond to the set S.

1.1.4 Organization

The rest of the chapter is organized as follows. We review basic properties of the DC programming and the DC functions in Section 1.2. We form a penalized least square problem with a generalized DC-penalty in Section 1.3. Under mild assumptions, we prove in Section 1.4 that a set of the d-stationary solutions are close to the ground truth. Furthermore, they are also support recovery consistent. We also extend our results to generalized loss functions, such as logistic loss, etc., in Section 1.5. We provide the connections among popular exiting algorithms in DC programming and statistics estimation with non-convex objective functions in Section 1.6. We finally conclude this work in Section 1.7. When possible, the technical proofs are relegated to the Appendix.

1.2 DC functions and related basic properties

In this section, we first provide the necessary background as well as a definition of the Difference-of-Convex (DC) functions, before proceeding to our formulation (Section 1.3) and the main results (Section 1.4 and 1.5). We present the definition of DC functions and its known properties in Section 1.2.1. The directional derivatives are reviewed in Section 1.2.2. The class of DC functions that we are particularly interested are reviewed in Section 1.2.3. We then define and study the directional stationarity (d-stationarity) that we focus on in this work in Section 1.2.4.

1.2.1 DC programming and DC functions

DC programming is pervasive nowadays in both optimization and statistics. The DC program has been introduced in the literature from 1950's [27]. The work in [28] gives a wealth of basic properties of the DC functions, which are the functions that are used in the objectives and constraints in the DC programming. In particular, the DC programming has been intensively studied in the field of optimization in the early 1980's [29, 30, 31, 32]. The following gives a formal definition for a DC function.

Definition 1.2.1. *A function, $p(x)$, is called a difference-of-convex (DC) function if we have*

$$p(x) = g(x) - h(x)$$

where both $g(x)$ and $h(x)$ are convex functions.

There are many known results regarding to DC functions and DC programming. We summarize these properties of DC programming from the literature in Appendix A.1.

1.2.2 Directional derivative

To enable our description, we define the *directional derivative* in the following.

Definition 1.2.2. *For a function $Q(\beta)$ that is defined on Ω where $\beta \in \Omega \subset \mathbb{R}$ or $\mathbb{R}^p$, for β_0, $\beta_1 \in \Omega$, the directional derivative at β_0 in the direction of $\beta_1 - \beta_0$ is defined as follows:*

$$Q'(\beta_0; \beta_1 - \beta_0) = \lim_{\tau \to 0+} \frac{Q(\beta_0 + \tau(\beta_1 - \beta_0)) - Q(\beta_0)}{\tau},$$

where $\tau \in \mathbb{R}_+$.

To compute the directional derivative, when a function $P(\beta)$ is differentiable in $\mathbb{R}^p$, the directional derivative with regard to β at β_0 is given below:

$$P'(\beta_0; \beta - \beta_0) = \langle \nabla P(\beta_0), \beta - \beta_0 \rangle,$$

where $\nabla P(\beta_0)$ is the gradient of the function $P(\beta)$ at β_0, and $\langle \cdot, \cdot \rangle$ represents the inner product. When the function $P(\beta)$ is non-differentiable however convex in $\mathbb{R}^p$, given its sub-gradient set $\partial P(\beta_0)$ at β_0, the directional derivative with regard to β at β_0 can be written as [33, Theorem 23.4]:

$$P'(\beta_0; \beta - \beta_0) = \max_{v \in \partial P(\beta_0)} \langle v, \beta - \beta_0 \rangle.$$

The recent works [34] and [1] discuss the pervasiveness of the existence of the DC functions as well as its relation to statistics. Specifically, they have the following results considering the pervasiveness of DC functions.

Lemma 1.2.1. *[34, Proposition 1] For any univariate continuous concave function p that is defined on $\mathbb{R}_+$, the composite function $\theta(|t|) = p(|t|)$ is Difference-of-Convex (DC) on $\mathbb{R}$ if and only if $p'(0; +)$, the directional derivative of $p(t)$ at 0, which can be written as follows,*

$$p'(0; +) = \lim_{\tau \to 0+} \frac{p(\tau) - p(0)}{\tau},$$

exists and is finite.

The above lemma leads to the realization that nearly all the folded-concave penalties [5, 6, 9, 10, 11] in the sparsity study nowadays belong to the DC family. We articulate details in the subsequent subsection.

1.2.3 Relation to statistics

Based on the definition of the DC functions (Definition 1.2.1), it has been realized (e.g., [35],[22], and [1]) that many well-studied penalties, such as SCAD, MCP, capped ℓ_1, transformed ℓ_1, logarithmic, can be written as DC functions. We articulate details in the following. Throughout this chapter, we consider penalties $P(\beta)$ that are separable in the (potentially multivariate) parameter β: $P(\beta) = \sum_{i=1}^{p} p(\beta_i)$, where $\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p$.

We argue that function $p(x)$ is a DC function for the popular existing penalties in the literature; that is, we have $p(x) = g(x) - h(x)$, where functions g and h are convex. More specifically, for the penalties that are of interests to us and are widely used in statistical inference, we always have $g(x) = |x|$ (or $g(x) = \lambda|x|$, when an algorithmic parameter λ is involved), however the function $h(x)$ varies per different penalties.

In the following, we describe how the popular penalty functions $p(x)$ mentioned previously can be decomposed as DC functions. For simplicity, without loss of generality, we set the tuning parameter as $\lambda = 1$.

1. In SCAD [5, 6], we have

$$p^{SCAD}(t) = |t| - h_\gamma^{SCAD}(t),$$

where

$$h_\gamma^{SCAD}(t) = \begin{cases} 0 & |t| \leq 1 \\ \frac{(|t|-1)^2}{2(\gamma-1)} & 1 \leq |t| \leq \gamma \\ |t| - \frac{(\gamma+1)}{2} & |t| \geq \gamma \end{cases}$$

and the function $h_\gamma^{SCAD}(t)$ can be verified to be convex on the positive real line and have a continuous first order derivative.

2. In MCP [9], we have

$$p^{MCP}(t) = |t| - h_\gamma^{MCP}(t),$$

where

$$h_\gamma^{MCP}(t) = \begin{cases} \frac{|t|^2}{2\gamma} & |t| \leq \gamma \\ |t| - \frac{\gamma}{2} & |t| \geq \gamma \end{cases}$$

and the function $h_\gamma^{MCP}(t)$ can be verified to be convex on the positive real line and have a continuous first order derivative.

3. In Capped ℓ_1 [10, 11], we have

$$p^{\text{capped } \ell_1}(t) = |t| - \max\left\{0, \frac{2t}{\gamma} - 1, -\frac{2t}{\gamma} - 1\right\},$$

where one can verify that both $|t|$ and $\max\left\{0, \frac{2t}{\gamma} - 1, -\frac{2t}{\gamma} - 1\right\}$ are convex functions of t.

4. In Transformed ℓ_1 [23, 24], we have

$$p^{\text{Transformed } \ell_1}(t) = \frac{a+1}{a}|t| - \left(\frac{a+1}{a}|t| - \frac{(a+1)|t|}{a+|t|}\right),$$

where one can verify that both $\frac{a+1}{a}|t|$ and $\left(\frac{a+1}{a}|t| - \frac{(a+1)|t|}{a+|t|}\right)$ are convex functions.

5. In Logarithmic [25], we have

$$p^{Log}(t) = \frac{1}{\epsilon}|t| - \left(\frac{|t|}{\epsilon} - \log(|t| + \epsilon) + \log\epsilon\right),$$

where ϵ is a given positive scalar; similarly, one can verify that both $\frac{\lambda}{\epsilon}|t|$ and $\left(\frac{|t|}{\epsilon} - \log(|t| + \epsilon) + \log$ are convex functions.

Table 1.1 summarizes the aforementioned DC decompositions. The penalty functions and their first order derivatives in terms of $|t|$ are plotted in Figure 1.1. One common point that most of the above penalties share is that their first order derivative goes to 0 as $|t| \to \infty$.

Although there are many other different DC decompositions, this one has the advantage of easy interpretation and correcting the penalty of LASSO, which in some sense, does a debiasing job for LASSO estimator (by choosing $h_\lambda(t)$ to be linear with slope λ when $|t|$ is large enough). It also shares common features with popular penalties in literature, like SCAD, MCP, capped ℓ_1 penalties where the penalty is close to or equal to ℓ_1 penalty when the solution is around the origin. Furthermore, the resulting penalty $p(x) = g(x) - h(x)$ is

Table 1.1: The DC decompositions of some well-known penalty functions in statistical inference. The first column contains the name of the methodology. The second column describes the penalty function. The last two columns present the corresponding two convex functional components (i.e., g and h) in the DC decomposition: $p(t) = g(t) - h(t)$.

Penalty	$p(t)$	$g(t)$	$h(t)$
ℓ_1	$\lvert t\rvert$	$\lvert t\rvert$	0
SCAD	$\int_0^{\lvert t\rvert} 1 \wedge (1 - \frac{x-1}{\gamma-1})_+ dx$	$\lvert t\rvert$	$\frac{(\lvert t\rvert-1)^2}{2(\gamma-1)} I\{1 < \lvert t\rvert < \gamma\}$ $+(\lvert t\rvert - \frac{\gamma+1}{2}) I\{\lvert t\rvert \geq \gamma\}$
MCP	$\int_0^{\lvert t\rvert} (1 - \frac{x}{\gamma})_+ dx$	$\lvert t\rvert$	$(\lvert t\rvert - \gamma/2) I\{\lvert t\rvert > \gamma\}$ $+\frac{t^2}{2\gamma} I\{\lvert t\rvert < \gamma\}$
Capped-ℓ_1	$\min\{\gamma/2, \lvert t\rvert\}$	$\lvert t\rvert$	$\max\{0, \frac{2t}{\gamma} - 1\}$
Transformed ℓ_1	$\frac{(a+1)\lvert t\rvert}{a+\lvert t\rvert}$	$\frac{a+1}{a}\lvert t\rvert$	$\frac{a+1}{a}\lvert t\rvert - \frac{(a+1)\lvert t\rvert}{a+\lvert t\rvert}$
Logarithmic	$\log(\lvert t\rvert + \epsilon) - \log \epsilon$	$\frac{\lvert t\rvert}{\epsilon}$	$\frac{\lvert t\rvert}{\epsilon} - \log(\lvert t\rvert + \epsilon) + \log \epsilon$

singular at 0, which makes it possible to achieve the condition of sparsity and continuity of the estimation [5]. The results in later sections can be applied to SCAD, MCP, capped-ℓ_1, and many others.

1.2.4 Directional stationary points

Another important definition in this chapter is the d(irectional)-stationary point, which is used to describe the set of stationary solutions we are interested in this chapter. We give the definition of the d-stationary point in the following.

Definition 1.2.3 (*d-stationary point*)**.** *A vector $\hat{\beta} \in \Omega$ is a d-stationary point to a function $Q(\beta)$ if the directional derivative, which is defined as*

$$Q'(\hat{\beta}; \beta - \hat{\beta}) = \lim_{\tau \to 0+} \frac{Q(\hat{\beta} + \tau(\beta - \hat{\beta})) - Q(\hat{\beta})}{\tau},$$

satisfies $Q'(\hat{\beta}; \beta - \hat{\beta}) \geq 0$ for all $\beta \in \Omega$.

We prove later that under some proper conditions on the penalty function as well as on the design matrix (which in some general cases are about the loss functions), a set of d-stationary solutions to the DC programming problem are $\sqrt{n}$ consistent estimators with a

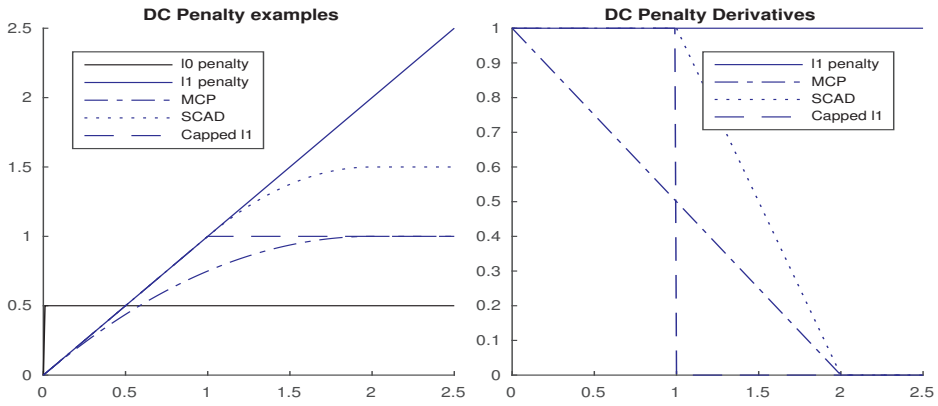

Figure 1.1: Examples of famous DC penalties and their derivatives in the literature: the ℓ_0 penalty is plotted in the solid black line; the ℓ_1 penalty function is plotted in the solid blue line; the MCP penalty is plotted in the dotted blue line; the MCP penalty is plotted in the dash-dot line; the Capped-ℓ_1 penalty is plotted in the dashed blue line.

high probability (Theorem 1.4.1). Under further cnditions, it also recovers the true support in the unknown parameter with a high probability (Theorem 1.4.3).

A motivation of choosing the directional stationary solutions (which are the directional stationary points in the corresponding optimization problem) rather than stationary solutions of other kinds, such as that of a critical point for DC programs, is provided in [1]. The authors argue that the directional stationary solutions are the sharpest kind among stationary solutions of other kinds in the sense a directional stationary solution must be stationary according to other definitions of stationarity. In the above sense, the d-stationary solutions possess minimizing properties that are not in general satisfied by stationary solutions of other kinds. We refer to the original chapter for a more detailed discussion.

1.3 Formulation and assumptions

In this section, we first give our detailed formulation in Section 1.3.1. We discuss the scale invariant properties of the formulation with some specific form of penalties in Section 1.3.2. We then list the assumptions on the penalty functions in Section 1.3.3, on the d-stationary solutions in Section 1.3.4, on the design matrix for our analytical study and corresponding justifications in Section 1.3.5.

1.3.1 Formulation

We present our problem formulation in the following. Recall that the widely-known SCAD [5] and MCP [9] choose their regularization term (i.e., the penalty function) as a function in the form, $\lambda|t| - h_\lambda(t)$, where the second term $h_\lambda(t)$ has a continuous first order derivative. Motivated by the above, we propose to analyze the following parameter estimation approach:

$$\min_{\beta\in\mathbb{R}^p} \frac{1}{2n}\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1 - h_\lambda(\beta), \tag{1.3.1}$$

where function $h_\lambda(\beta)$ is assumed to be convex and the model is specified in (1.1.1). As we have shown in Table 1.1, popular non-convex penalties in the literature can all be expressed in DC form. In our formulation, following the approaches in the main stream methodology, we focus on separable penalty, that is we have

$$P(\beta) = \lambda\|\beta\|_1 - h_\lambda(\beta) = \sum_{i=1}^{p} \lambda|\beta_i| - h_\lambda(\beta_i),$$

for $\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p$. Note that based on the context, function $h_\lambda(\cdot)$ can take both univariate and multivariate inputs. In our formulation, the univariate function $h_\lambda(\cdot)$ is assumed to be convex. Its properties are further specified later.

1.3.2 Scale-invariant property

In real world of processing data, programmers always perform rescaling on the raw data set. We can make our formulation scale-invariant by assuming the following format of the penalty function.

Assumption 1.3.1. $p_\lambda(t) = \lambda^2 p(\frac{t}{\lambda})$.

Suppose we scale the model in (1.1.1) by a scalar c, which leads to the following model:

$$cY = X(c\beta^*) + c\epsilon,$$

Let $F(\beta, \lambda) = \frac{1}{2n}\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1 - h_\lambda(\beta)$ denote the objective function, corresponding to (1.1.1). Let $F(c\beta, c\lambda) = \frac{1}{2n}\|cY - X(c\beta)\|_2^2 + c\lambda\|c\beta\|_1 - h_{c\lambda}(c\beta)$ denote the objective function, corresponding to the scaled model. One can easily verify that for any given positive scalar c,

$$\min_{\beta \in \mathbb{R}^p} F(c\beta, c\lambda), \tag{1.3.2}$$

is equivalent to the original problem of $\min_{\beta \in \mathbb{R}^p} F(\beta, \lambda)$ in (1.3.1) with scale free penalties such as SCAD, ℓ_1, MCP, capped-ℓ_1, which have the common form stated in Assumption 1.3.1. One can verify that most of the functions in Table 1.1 satisfy this scale free condition except the logarithmic function and the transformed ℓ_1. However, according to the unification DCA in Section 1.6, where in each iteration, by using only the linear approximation, we solve a re-weighted LASSO problem, which is scale invariant.

1.3.3 Assumptions on $h_\lambda(\cdot)$

We present the assumptions that we need in the analytical study in this section. Recall that our penalty function has the form $P(\beta) = \lambda\|\beta\|_1 - h_\lambda(\beta)$. Notice that the first term of the DC penalty is always the ℓ_1 function. We specify our assumptions on the univariate function $h_\lambda(\beta)$, for $\beta \in \mathbb{R}$. We also require the regularity conditions on the design matrix X, which is articulated later.

The following assumptions are utilized in our analysis. We briefly discuss the assumptions and argue that our assumptions are equivalent or weaker to conditions in most existing work.

Assumption 1.3.2. *We have* $\sup_{t \in \mathbb{R}} |h'_\lambda(t)| \leq \lambda$.

Assumption 1.3.3. *We have* $h_\lambda(t)$ *is symmetric about 0.*

Both Assumption 1.3.2 on the non-negativity of the penalty and Assumption 1.3.3 on the symmetry of the penalty function are standard assumptions in the literature. Assumption 1.3.2 makes sure that the penalty function $P_\lambda(\beta_i)$ is nonnegative. In fact, we can even

relax this condition to $\sup_{t\in\mathbb{R}} |h'_\lambda(t)| \leq \lambda$ as long as the first order derivative of the function $h_\lambda(t)$ is uniformly bounded in the real line.

Assumption 1.3.4. *$h'_\lambda(t)$ is monotonically increasing and there exist two nonnegative constants $\eta^- \geq \eta^+ \geq 0$ such that for any $t_2 > t_1$:*

$$0 \leq \eta^+ \leq \frac{h'_\lambda(t_2) - h'_\lambda(t_1)}{t_2 - t_1} \leq \eta^- \tag{1.3.3}$$

Regarding Assumption 1.3.4, the lower bound η^+ of the convexity of the function $h_\lambda(t)$ is usually assumed to be 0 in other works, such as in the SCAD and the MCP. The upper-bound of the convexity η^- is used to control the convexity of the function $h_\lambda(t)$. If $h_\lambda(t)$ has a "lot" of convexity, we are not able to have the Restricted Strong Convexity of the objective function later. On the other hand, this can be regarded as requiring the first order derivative of $h_\lambda(t)$ to be continuous. The continuity assumption together with Assumption 1.3.2 and 1.3.6 ensure that Assumption 1.3.4 holds.

Assumption 1.3.5. *We have $h_\lambda(0) = h'_\lambda(0) = 0$.*

Assumption 1.3.5 is utilized to ensure the soft thresholding property of the penalty function [5], recalling that the singularity of the whole penalty function at 0.

Assumption 1.3.6. *For some positive ζ, we have $h'_\lambda(t) = \lambda$ for all $|t| \geq \zeta$.*

Assumption 1.3.6 is based on the fact [5] that making sure $h'_\lambda(t) = \lambda$ for t positive enough and $h'_\lambda(t) = -\lambda$ for t negative enough help producing an unbiased estimator. Recall that one of the main reasons of considering a generalized version of the LASSO method is the bias of the estimation from LASSO.

Below, we make a table of the penalties discussed in Table 1.1 and presents the decomposition to $\lambda|t| - h_\lambda(t)$. We also listed the properties that each of the $h_\lambda(t)$ holds.

From Table **??**, we can see that, SCAD and MCP penalty class satisfy all of the assumptions. While Capped-ℓ_1 has discontinuous first order derivative, which violates the

Table 1.2: The penalties in the sparse estimation literature and their properties with respect to our assumptions. The first column gives the name of the methods. The second column presents the h-function, which is the second component in the DC decomposition ($p = g - h$) of the corresponding penalty function. The third column contains their first derivatives on the positive axe. This is to verify Assumption 1.3.2. The fourth column computes for the quantities that are raised in Assumption 1.3.4. The last column summarizes the assumptions that the corresponding penalty satisfies.

Penalty	$h(t)$	sgn(t)$h'(t)$	Convexity measure	Assumptions
ℓ_1	0	0	0	All except 1.3.6
Capped-ℓ_1	$\max\{0, \frac{2t}{\gamma} - 1\}$	$\frac{2}{\gamma} I\{\lvert t\rvert > \gamma/2\}$	∞	All except 1.3.4
MCP	$(\lvert t\rvert - \gamma/2) I\{\lvert t\rvert > \gamma\}$ $+\frac{t^2}{2\gamma} I\{\lvert t\rvert < \gamma\}$	$\min\{\frac{\lvert t\rvert}{\gamma}, 1\}$	γ^{-1}	All
SCAD	$\frac{(\lvert t\rvert-1)^2}{2(\gamma-1)} I\{1 < \lvert t\rvert < \gamma\}$ $+(\lvert t\rvert - \frac{\gamma+1}{2}) I\{\lvert t\rvert \geq \gamma\}$	$\frac{\lvert t\rvert-1}{\gamma-1} I\{1 < \lvert t\rvert < \gamma\}$ $+I\{\lvert t\rvert \geq \gamma\}$	$(\gamma - 1)^{-1}$	All
Transformed ℓ_1	$\lvert t\rvert - \frac{a\lvert t\rvert}{a+\lvert t\rvert}$	$\frac{(a+\lvert t\rvert)^2 - a^2}{(a+\lvert t\rvert)^2}$	$\frac{2}{a}$	All except 1.3.6
Logarithmic	$\lvert t\rvert - \log(\lvert t\rvert + 1)$	$\frac{\lvert t\rvert}{\lvert t\rvert+1}$	1	All except 1.3.6

Assumption 1.3.4. In order to extend the theories in this work to Capped-ℓ_1 penalty, performing smoothing around the non-differentiable point is enough. We re-scaled the linear term in Transformed ℓ_1 to match the assumptions. ϵ in Logarithmic penalty is chosen to be 1 in the Table **??**.

1.3.4 Assumptions on the d-stationary solution

Besides the assumptions on the penalty functions, we also list the assumptions necessary for the loss function. These are about the design matrix in case of linear model with the least square loss function.

Assumption 1.3.7. *Let β^* be the unknown true parameter, $\hat{\beta}$ be the d-stationary solution to problem (1.3.1), which satisfies the following condition:*

$$\frac{1}{n} X_j^T X(\beta - \hat{\beta}) \textit{sign}(\hat{\beta}_j) \geq c\lambda, \textit{ for all } j \in S^c, \beta = \beta^* \textit{ with } c \in (0, 1).$$

Remark 1.3.1. *The above Assumption 1.3.7 is no stronger than the assumptions used in LASSO estimator [4]. We show below that in the proof of LASSO estimator, it corresponds to when* $c = \frac{1}{2}$. *Let* $\hat{\beta}^{lasso} = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{2n}\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1$. *Recall that in [4], by the First Order Condition (FOC) at* $\hat{\beta}^{lasso}$, *we have*

$$-\frac{1}{n}X^T(Y - X\hat{\beta}^{lasso}) + \lambda\partial\|\hat{\beta}^{lasso}\|_1 = 0,$$

where $\partial\|\hat{\beta}^{lasso}\|_1$ *is a subgradient at* $\hat{\beta}^{lasso}$ *for* $\|\beta\|_1$. *Multiply by* $(\beta^* - \hat{\beta}^{lasso})^T$ *on both side, we have*

$$-\frac{1}{n}(\beta^* - \hat{\beta}^{lasso})^T X^T(Y - X\hat{\beta}^{lasso}) + \lambda(\beta^* - \hat{\beta}^{lasso})^T\partial\|\hat{\beta}^{lasso}\|_1 = 0.$$

Since

$$\begin{aligned}
&(\beta^* - \hat{\beta}^{lasso})^T\partial\|\hat{\beta}^{lasso}\|_1 \\
=&(\beta^* - \hat{\beta}^{lasso})_S^T\partial\|\hat{\beta}_S^{lasso}\|_1 + (\beta^* - \hat{\beta}^{lasso})_{S^c}^T\partial\|\hat{\beta}_{S^c}^{lasso}\|_1 \\
=&(\beta^* - \hat{\beta}^{lasso})_S^T\partial\|\hat{\beta}_S^{lasso}\|_1 - \|\hat{\beta}_{S^c}^{lasso}\|_1
\end{aligned} \tag{1.3.4}$$

Plugging into the FOC, we have

$$\begin{aligned}
&\frac{1}{n}(\beta^* - \hat{\beta}^{lasso})^T X^T(Y - X\hat{\beta}^{lasso}) \\
=&\frac{1}{n}(\beta^* - \hat{\beta}^{lasso})^T X^T X(\beta^* - \hat{\beta}^{lasso}) + \frac{1}{n}(\beta^* - \hat{\beta}^{lasso})^T X^T\epsilon \\
=&\frac{1}{n}(\beta^* - \hat{\beta}^{lasso})^T X^T X(\beta^* - \hat{\beta}^{lasso}) + \frac{1}{n}(\beta^* - \hat{\beta}^{lasso})_S^T X_S^T\epsilon \\
&+ \frac{1}{n}(\beta^* - \hat{\beta}^{lasso})_{S^c}^T X_{S^c}^T\epsilon \\
=&\frac{1}{n}(\beta^* - \hat{\beta}^{lasso})^T X^T X(\beta^* - \hat{\beta}^{lasso}) + \frac{1}{n}(\beta^* - \hat{\beta}^{lasso})_S^T X_S^T\epsilon \\
&- \sum_{S^c}\frac{1}{n}|\hat{\beta}^{lasso}|_i X_i^T\epsilon \boldsymbol{sign}(\hat{\beta}_i^{lasso}) \\
=&\lambda(\beta^* - \hat{\beta}^{lasso})_S^T\partial\|\hat{\beta}_S^{lasso}\|_1 - \lambda\|\hat{\beta}_{S^c}^{lasso}\|_1
\end{aligned} \tag{1.3.5}$$

where in the first equality, we plugged in $Y = X\beta^* + \epsilon$. *If we have the condition that* $\frac{1}{n}X_i^T X(\beta^* - \hat{\beta}^{lasso})$*sign*$(\hat{\beta}_i^{lasso}) > c\lambda$ *(*$c = \frac{1}{2}$ *in LASSO) for all* $i \notin S$, *we have*

$$\begin{aligned} &\frac{1}{n}(\beta^* - \hat{\beta}^{lasso})^T X^T X(\beta^* - \hat{\beta}^{lasso}) \\ &\leq \frac{3}{2}\lambda\|\hat{\beta}_S^{lasso}\|_1 - c\lambda\|\hat{\beta}_{S^c}^{lasso}\|_1 \end{aligned} \tag{1.3.6}$$

with high probability. Similarly, we made Assumption 1.3.7 in the generalized penalized regression to ensure good property of the solution.

Remark 1.3.2. *Since the condition in Assumption 1.3.7 cannot be verified directly, in real data analysis, we can use the following checkable conditions instead:*

$$\frac{1}{n}X_j^T(Y - X\hat{\beta})\textit{sign}(\hat{\beta}_j) \geq c\lambda, \textit{ for all } j \in S^c \textit{ such that } \hat{\beta}_j \neq 0, \beta = \beta^*,$$

where c *is defined in Assumption 1.3.7. If the above holds, the Assumption 1.3.7 hold with high probability using similar argument of sub-Gaussian random variables as in the proof of Theorem 1.4.2.*

1.3.5 Assumptions on the design matrix X

Definition 1.3.1. *The restricted strong convexity (RSC) condition on model matrix* X *with respect to* $\mathcal{C}$ *is the following, there exists some constant* $\gamma > 0$ *such that:*

$$\frac{\frac{1}{N}\nu X^T X\nu}{\|\nu\|_2^2} \geq \gamma \textit{ for all nonzero } \nu \in \mathcal{C}$$

here γ is called the restricted eigenvalue bound with regard to $\mathcal{C}$.

Assumption 1.3.8. *Denote* $C_{\mathcal{S}}$ *by the diagonal matrix with* $\{c_i, i \in \mathcal{S}\}$, $C_{\mathcal{S}^C}$ *by the diagonal matrix with* $\{c_i, i \in \mathcal{S}^C\}$, *the restricted eigenvalues (RE) condition holds on the*

following set with some positive c defined in Assumption 1.3.7:

$$\mathcal{C} = \left\{ \nu \in \mathbb{R}^p \,\middle|\, \|C_{\mathcal{S}^C} \cdot \nu_{\mathcal{S}^C}\|_1 \leq \frac{5}{2c} \|C_{\mathcal{S}} \cdot \nu_{\mathcal{S}}\|_1 \right\},$$

where $\cdot$ indicates the matrix-vector multiplication.

We have $\mathcal{C} \subset \mathbb{R}^p$ strictly since it is of the form of cone.

The RSC (Assumption 1.3.8) is a standard assumption in the literature for proving the consistency results of regularized high-dimensional sparse estimation problems.

1.4 Consistency results for some d-stationary solutions

We prove our main results in this section. The non-asymptotic upper bound for estimation errors is derived in Section 1.4.1. As a corollary, we provide the upper bound for prediction errors as a byproduct in Section 1.4.2. We provide the results regarding asymptotic consistency of the estimation in Section 1.4.3. The asymptotic consistency in support recovery is discussed in Section 1.4.4.

1.4.1 Non-asymptotic upper bound for estimation errors

In this section, we present our results on the non-asymptotic upper bound for the estimation error. We mainly use the assumptions on the model matrix to prove that the difference between the ground truth β^* and the d-stationary solution $\hat{\beta}_n^\lambda$ will be in a cone-like set, where we have the restricted strong convexity (RSC) assumption hold (as defined in Assumption 1.3.8). Without Assumption 1.3.4 on the continuity of the first order derivative on the function $h_\lambda(t)$, we will be able to obtain the upper-bound of the ℓ_2 distance between the ground truth and the d-stationary estimation.

Theorem 1.4.1. *Suppose $h_\lambda(t)$ satisfies Assumptions 1.3.2, 1.3.3, 1.3.5, design matrix X satisfies the restricted strong convexity with respect to $\mathcal{C}$, which is defined in Assumption 1.3.8 with $c_i = 1$ for $i = 1, \cdots, p$, with $\lambda \geq \frac{2\|X^T\epsilon\|_\infty}{n}$. If we further assume Assumption*

1.3.7 holds at the d-stationary solution $\hat{\beta}_n^\lambda$, we will have the upper bound for estimation error on β^* with the d-stationary estimation $\hat{\beta}_n^\lambda$ as $n \to \infty$:

$$\|\hat{\beta}_n^\lambda - \beta^*\|_2 \leq \frac{5}{2\gamma}\lambda\sqrt{\|\beta^*\|_0}$$

The proof of the above Theorem 1.4.1 is delayed to Appendix A.2.1. The results in Theorem 1.4.1 suggest that the d-stationary solution to Problem (1.3.1), under mild conditions, will be able to retrieve the information in the unknown parameter β^* with error bounded within the order of $O(\frac{\lambda\sqrt{|S|}}{\gamma})$, which is optimal.

1.4.2 Non-asymptotic upper bound for prediction errors

From the proof of Theorem 1.4.1, we will be able to further give the upper bound for the prediction error below.

Corollary 1.4.1. *Under the assumptions of Theorem 1.4.1, we can further get the upper-bound for the prediction error as:*

$$\left\|\frac{1}{\sqrt{n}}X(\beta^* - \hat{\beta}_n^\lambda)\right\|_2^2 \leq (\frac{5}{2}\lambda)^2\frac{1}{\gamma}|S|$$

The proof of Corollary 1.4.1 is straight forward according to the proof in Theorem 1.4.1 and is postponed to Appendix A.2.2.

1.4.3 Asymptotic convergence rate

If we further assume that the errors are independent sub-Gaussian distributed, we will be able to bound the estimation error in the order of $\sqrt{\frac{|S|\log p}{n}}$ with high probability.

Corollary 1.4.2. *Under the assumptions of Theorem 1.4.1, if we further assume that the errors are from independent sub-Gaussian with variance parameter σ^2 and mean 0, we will*

have the following hold with probability at least $1 - 2\exp\left(-\frac{\tau-2}{2}\log p\right)$

$$\|\hat{\beta}_n^{\lambda} - \beta^*\|_2 \lesssim \frac{5}{\gamma}\sigma\sqrt{\frac{\tau|S|\log p}{n}}.$$

We provide the proof of Corollary 1.4.2 in Appendix A.2.3.

Remark 1.4.1. *All the results above are considering the problem in (1.3.1). However, the conclusions will still hold for the constrained version of problem (1.3.1) as long as the assumptions are satisfied. The results in this work assumes the d-stationary solution that satisfies our assumptions exists. We will justify the existence of the d-stationary solution satisfying our assumptions in Section 1.5.1.*

1.4.4 Support recovery

In this section, we will first provide the KKT conditions for d-stationary solutions, which says that the d-stationary condition in our work is equivalent to the first order condition in case of no constraints. Then we prove the Restricted Strong Convexity for the objective function in Problem (1.3.1) under some regularity conditions. By usage of the oracle estimator defined later in Problem (1.4.3), we will be able to prove the support recovery consistency of some of the d-stationary solutions to Problem (1.3.1).

Lemma 1.4.1. *Let* $F(\beta) = L(\beta) + g(\beta) - h(\beta)$*, where* $L(\beta)$*,* $g(\beta)$*,* $h(\beta)$ *are convex with* $\beta \in \mathbb{R}^p$*. Further assume that* $L(\beta)$ *and* $h(\beta)$ *have continuous first order derivative,* $g(\beta) = \|\beta\|_1$*. Let* β_0 *be a d(irectional)-stationary solution to* $F(\beta)$*, we have the following first order condition (FOC) hold at* β_0*. We will be able to get the following equivalent condition:* β_0 *is a d(irectional)-stationary solution to* $F(\beta)$ *if and only if there exists some* $z \in \partial g(\beta_0)$*, where* $\partial g(\beta_0)$ *is the set of subgradient of* $g(\beta)$ *at* β_0*, such that:*

$$\nabla L(\beta_0) + z - \nabla h(\beta_0) = 0, \tag{1.4.1}$$

where $\nabla L(\beta_0)$, $\nabla h(\beta_0)$ is the gradient of L, h at β_0.

The above Lemma 1.4.1 states the equivalence between d-stationary solution and first order condition (FOC) in the unconstrained case. While in constrained case, this does not necessarily hold. From the proof Lemma 1.4.1, we can derive similar conditions for "local maximals" for $\tilde{\beta}$. We obtain that as long as $\min_{i=1}^{p}\{\tilde{\beta}_i\} = 0$, it will only satisfy the condition for "local" minimals and thus be a d-stationary solution. Thus, in order to find the d-stationary solution, we only need to find a β_0 such that, there exists a vector $z \in \partial\|\beta_0\|_1$, the subgradient of function $\|\beta\|_1$ at $\beta = \beta_0$, such that $\nabla L(\beta_0) - \nabla h(\beta_0) + z = 0$. Furthermore, if $\min_{i=1}^{p}\{z_i\} < 1$, which is known as the strict dual feasibility condition [4], it will be satisfying the condition for "local" minimals.

Remark 1.4.2. *Generally, a d-stationary solution is not necessarily local minimal. For example, for a differentiable function $f(x, y) = x^2 - y^2$, where both the function $g(x, y)$ and $h(x, y)$ are differentiable (slightly different from the above situation in Lemma 1.4.1), at a saddle point $(0, 0)$, which is stationary with $\mathbf{0}$ gradient, the directional derivative at this point will all be 0, which makes it a d-stationary solution however not a local minimal. Another example would be $f(x, y) = |x| - y^2$ at the saddle point $(0, 0)$.*

Remark 1.4.3. *The necessary condition to be a local minimal is being a d-stationary point in the feasible region.*

The following Lemma shows the RSC of the Problem (1.3.1).

Lemma 1.4.2. *Under Assumption 1.3.8 with h_λ satisfying Assumptions 1.3.2, 1.3.3, 1.3.4, 1.3.5, let β_1, $\beta_2 \in \mathbb{R}^p$ such that $\nu = \beta_1 - \beta_2 \in \mathcal{C}$, where $\mathcal{C}$ is defined in Assumption 1.3.8. Then $f_\lambda(\beta) = \frac{1}{2n}\|Y - X\beta\|_2^2 - h_\lambda(\beta)$ will satisfy the restricted strong convexity given that $\gamma > \eta^-$:*

$$f_\lambda(\beta_2) \geq f_\lambda(\beta_1) + \nabla f_\lambda(\beta_1)^T(\beta_2 - \beta_1) + \frac{\gamma - \eta^-}{2}\|\beta_2 - \beta_1\|_2^2 \tag{1.4.2}$$

The proof can be found in Appendix A.2.5

Oracle estimator

The oracle estimator is defined as follows:

$$\beta^O = \arg \min_{\beta \in \mathbb{R}^p, \beta_{S^c}=0} \frac{1}{2n} \|Y - X\beta\|_2^2. \tag{1.4.3}$$

The oracle estimator is obtained as if there is an oracle telling the true support of the underlying unknown estimator. According to the definition of oracle estimator β^O, we are able to provide the following ℓ_∞ error bound between β^O and β^*. We also demonstrate that β^O is a d-stationary solution to the DC-penalized Problem (1.3.1), which we are interested in this chapter. The following Theorem 1.4.2 and Lemma 1.4.3 also appeared in the work by Wang et al. [22]

Theorem 1.4.2. *Under Assumption 1.3.8, the oracle estimator is the unique global minimizer of (1.4.3). If the noise is independent sub-Gaussian with variance parameter σ^2, the oracle estimator will satisfy the following ℓ_∞ error bound with high probability.*

$$\|\beta^O - \beta^*\|_\infty \leq C\sigma\sqrt{2/\gamma}\sqrt{\frac{\log s}{n}}.$$

The proof is in Appendix A.2.6.

Lemma 1.4.3. *Under Assumption 1.3.8 with h_λ satisfying Assumptions 1.3.2, 1.3.3, 1.3.4, 1.3.5, 1.3.6, let β^O be the aforementioned oracle estimator. Assume further that for the ground truth β^*, we have $\min_{i=1}^p |\beta_i^*| > 2\zeta$, for $\zeta > 0$. There exists a subgradient $\xi^O \in \partial\|\beta^O\|_1$ (where $\partial\|\beta^O\|_1$ stands for the subgradient of function $\|\beta\|_1$ at $\beta = \beta^O$) such that for any $\beta \in \mathbb{R}^p$:*

$$(\beta - \beta^O)^T \left(\nabla f_\lambda(\beta^O) + \lambda\xi^O\right) \geq 0 \tag{1.4.4}$$

The above Lemma 1.4.3 assumes that the penalty on the parameters will be a constant when the parameters are large. As it requires Assumption 1.3.6, the result is not applicable

for transformed ℓ_1 and logarithmic penalties. We postpone the proof to Appendix A.2.7.

Lemma 1.4.4. *Under the assumptions in Lemma 1.4.3, let $\hat{\beta}$ be a d-stationary solution to (1.3.1) satisfying Assumption 1.3.7, and β^O be the oracle estimator. The following will hold with large probability:*

$$\nu = \hat{\beta} - \beta^O \in \mathcal{C}. \tag{1.4.5}$$

The proof is in Appendix A.2.8. Based on the previous results of the oracle estimator β^O and properties of the d-stationary estimator $\hat{\beta}$, we will now be able to prove the support recovery consistency for our generalized DC-penalized model.

Theorem 1.4.3. *Under the conditions of Lemma 1.4.4, we will have $supp(\hat{\beta}) = supp(\beta^O) = supp(\beta^*)$ with high probability.*

The proof is provided in Appendix A.2.9. We prove the support recovery consistency for a set of d-stationary solutions, it implies that a set of the convergence points (satisfying Assumption 1.3.7) from the DCA will converge to the oracle estimator which is unique. In the work from Wang et al, they prove that the convergence point from each stage of the specific algorithm converges to the oracle estimator in the linear model setting. The above results also inform us how we should choose the penalty function such that the d-stationary solution will be support recovery consistent. The penalty needs to be a constant when the parameter gets larger (Assumption 1.3.6), so that the resulting oracle estimator will be a d-stationary solution to the original Problem (1.3.1). Assumption 1.3.4 is necessary for the restricted strong convexity in $\mathcal{C}$.

1.5 DC penalty with generalized loss functions

In the previous section, we mainly focus on the linear model scenario. Most of the analysis can be readily extended to generalized loss functions such as logistic loss function, etc. Below, we will present the formulation of DC penalized likelihood and provide the statistical analysis regarding to the d-stationary solutions.

We begin with a brief review on the exponential family. Exponential family is a family of distributions with the probability density proportional to $P(Y|X,\beta^*) \propto \exp\{\frac{YX^T\beta^* - \psi(X^T\beta^*)}{c(\sigma)}\}$, where $c(\sigma)$ is a scaling parameter and $\psi(\cdot)$ is the cumulant function. According to [36], one standard property of exponential family is

$$\psi'(X^T\beta^*) = \mathbb{E}[Y|X,\beta^*,\sigma].$$

Given that $\psi(\cdot)$ is a univariate convex function, let $L(\beta) = \psi(X^T\beta) - YX^T\beta$ be the negative log likelihood function, $L_n(\beta) = \frac{1}{n}\sum_{i=1}^n(\psi(X_i^T\beta) - Y_iX_i^T\beta)$ be the sample average of the negative log likelihood function, one can easily check that $\mathbb{E}[\nabla L(\beta^*)] = 0$ and $\nabla^2 L_n(\beta) \geq 0$. This implies that $L_n(\beta)$ is convex. In the following, we might omit the subscript n in the expression of $L_n(\beta)$ where no confusion will rise. In order to estimate the sparse ground truth β^*, we will solve the following DC penalized optimization problem:

$$\min_{\beta\in\mathbb{R}^p} \frac{1}{n}\sum_{i=1}^n(\psi(X_i^T\beta) - Y_iX_i^T\beta) + \lambda\|\beta\|_1 - h_\lambda(\beta), \tag{1.5.1}$$

Below, we will state the assumptions on the generalized loss functions, which enable us to provide the analysis that the error between the d-stationary solution $\hat{\beta}$ and the ground truth β^* is of the order $\mathbb{O}(\frac{17\lambda}{8(\gamma-\eta^-)}\sqrt{|S|})$.

Assumption 1.5.1. *Let β^* represent the ground truth of the unknown parameter, $L(\beta)$ be the negative log likelihood function. Assume that the infinity norm of the gradient of the loss function at the ground truth $\|\nabla L(\beta^*)\|_\infty \leq \frac{\lambda}{8}$.*

Assumption 1.5.2. *Let β^* be the unknown true parameter, $\hat{\beta}$ be the d-stationary solution to problem (1.3.1), which satisfies the following condition:*

$$\|\nabla h_\lambda(\hat{\beta}_{S^c})\|_\infty \leq (1-c)\lambda, \text{ with } c \in (0,1).$$

Assumption 1.5.3. *Let β^* represent the ground truth of the unknown parameter, $L(\beta)$ be*

the negative log-likelihood function. Assume that the following restricted strong convexity holds on the set $\mathcal{C}$,

$$\begin{aligned} \mathcal{C} &= \left\{ \nu \in \mathbb{R}^p \,\middle|\, \|C_{\mathcal{S}^C}\nu_{\mathcal{S}^C}\|_1 \leq \frac{4+c}{c}\|C_{\mathcal{S}}\nu_{\mathcal{S}}\|_1 \right\}, \\ L(\beta_1) &\geq L(\beta_2) + \nabla L(\beta_2)^T(\beta_1 - \beta_2) + \frac{\gamma}{2}\|\beta_1 - \beta_2\|_2^2, \end{aligned}$$

for any β_1 and β_2 such that $\beta_1 - \beta_2 \in \mathcal{C}$.

Theorem 1.5.1. *Let $\hat{\beta}$ be the d-stationary solution to the penalized loss function in (1.5.1). Suppose $h_\lambda(t)$ satisfies Assumptions 1.3.2, 1.3.3, 1.3.4, 1.3.5, assume further that Assumptions 1.5.1, 1.5.3 and 1.5.2 hold, if $\gamma > \eta^-$,*

$$\left\| \frac{1}{n}\sum_{i=1}^{n}(\psi'(X_i^T\beta^*)X_i - Y_iX_i) \right\|_\infty \leq \frac{c}{2},$$

we will have the following upper-bound for the estimation error of the d-stationary solution

$$\|\beta^* - \hat{\beta}\|_2 \leq \frac{17\lambda}{8(\gamma - \eta^-)}\sqrt{|S|}$$

The proof is provided in Appendix A.3.1.

1.5.1 Existence of d-stationary solution

In this section, we will show the existence of the d-stationary solutions we studied above. It is easy to see that in the linear regression setting with square loss, the oracle estimator is a d-stationary solution under suitable conditions we stated in Lemma 1.4.3. For general settings with generalized loss functions, let $r_0 > 0$ be such that $h'_\lambda(r_0) = (1-c)\lambda$, consider the following constrained problem:

$$\min_{\|\beta - \beta^*\|_2 \leq r} L_n(\beta) + \lambda\|\beta\|_1 - h_\lambda(\beta), \tag{1.5.2}$$

where $r = c\lambda\sqrt{|S|} \wedge r_0$. It is straightforward to check that the sationary solutions to Problem (1.5.2) satisfies all the assumptions of the d-stationary solution studied in Section 1.4 and Section 1.5, which verifies the existence of the wanted d-stationary solutions.

1.6 Numerical Approach to Find the d-stationary Points

In this section, we will review the efficient algorithms in the DC-literature, for finding the local optima in the statistics and optimization areas. This provides a comprehensive summary on solving DC programming. Up to our knowledge, the most classic algorithm is the Difference-of-Convex Algorithm (DCA) studied in [31, 29, 31, 37], which iterates between the primal problem and the dual problem to find the local minima. Given the DC problem below:

$$\min_{x\in\mathbb{R}^n} f(x) = g(x) - h(x), \tag{1.6.1}$$

where $g(\cdot)$ and $h(\cdot)$ are convex functions. For a function $g : \mathbb{R}^n \to \mathbb{R}$, let $g^*(y)$ be its convex conjugate function, which is defined as $g^*(y) = \sup\{x^T y - g(x) : x \in \mathbb{R}^n\}$. We have

$$\begin{aligned}
&\inf_{x\in\mathbb{R}^n} f(x) = g(x) - h(x) \\
=&\inf_x\{g(x) - \sup_y\{x^T y - h^*(y)\}\} \\
=&\inf_x\{\inf_y\{g(x) + h^*(y) - x^T y\}\} \\
=&\inf_y\{-\sup_x\{x^T y - g(x)\} + h^*(y)\} \\
=&\inf_y\{h^*(y) - g^*(y)\}.
\end{aligned} \tag{1.6.2}$$

Thus, by iterating between the primal and the dual problems, the DCA will converge to a d-stationary solution. Below shows the DCA.

According to [38] in Section 2.5, DCA has linear convergence rate for general DC programmings. While in the statistics literature, Local Linear Approximation (LLA) in

1: Choose the initial x_0
2: *loop*:
3: **for** $k \in \mathbb{N}$ **do**
4: Choose $y_k \in \partial h(x_k)$.
5: Choose $x_{k+1} \in \partial g^*(y_k)$.
6: **if** $(\min\{|(x_{k+1} - x_k)_i|, |\frac{(x_{k+1}-x_k)_i}{(x_k)_i}|\} \leq \delta)$ **then**
7: **return** x_{k+1}
8: **end if**
9: **end for**

Algorithm 1: Difference-of-Convex Algorithm (DCA)

[26] is widely used for solving regularized estimation problems with non-convex penalties. The update at each iteration takes the LLA of the penalty function:

$$x^{k+1} = \arg\min\{g(x) - \partial h(x_k)^T x\},$$

which is exactly the same procedure as shown in the Algorithm **??**.

In the setting of this chapter, the objective is defined in (1.3.1), where we are minimizing the objective function over all $\beta \in \mathbb{R}^p$ with the first part of the DC function as $g(\beta) = L_n(\beta) + \lambda\|\beta\|_1$, and the second part of DC function $h(\beta) = h_\lambda(\beta)$. The DCA can be simplified to Local Linear Approximation (LLA) in the general case as in [26], the detailed procedures can be found in [37]. Specifically, if $h(x)$ is differentiable, we will have the following equivalent algorithm as DCA:

1: Choose the initial β_0
2: *loop*:
3: **for** $k \in \mathbb{N}$ **do**
4: Choose $z_k \in \nabla h(\beta_k)$.
5: $\beta_{k+1} = \arg\min L_n(\beta) + \lambda\|\beta\|_1 - \langle \beta, \nabla h(\beta_k)\rangle$.
6: **if** $(\min\{|(\beta_{k+1} - \beta_k)_i|, |\frac{(\beta_{k+1}-\beta_k)_i}{(\beta_k)_i}|\} \leq \delta)$ **then**
7: **return** β_{k+1}
8: **end if**
9: **end for**

Algorithm 2: DCA (LLA)

According to [39], DCA is exactly the formulation of Convex Concave Procedure

(CCCP), which is also discussed in [40]. Thus, under proper conditions, all results in [39] can be applied to the problem studied here. Since our formulation (1.3.1) is a special form of the model considered in [40], which adopts the classical algorithm DCA (Difference-of-Convex Algorithm) in [31, 29, 31, 37] and solves a strictly convex problem at each iteration, it is guaranteed to converge quickly to a d-stationary solution. Since the penalty is a function of the absolute value of the estimator, one minor change to the above algorithm would be solving the following transformed optimization problem within each iteration:

$$\beta_{k+1} = \arg\min L_n(\beta) + \sum_{i=1}^{p} (\lambda - h'(|\beta_{ki}|))|\beta_i|, \tag{1.6.3}$$

which is exactly the formulation of weighted LASSO estimator and can be solved efficient using the LARS algorithm in [41].

Lemma 1.6.1. *By updating the parameter β as in Procedure 1.6.3, the objective function $F(\beta)$ defined in 1.3.1 is monotonically decreasing.*

In the one-step LLA procedure [26], the authors prove that starting from the maximum likelihood estimator (MLE), after one step of the LLA update, the resulting estimator is consistent when SCAD penalty function is used. While in [42], they prove that from the LASSO initialization, with high probability that the LLA converges to the oracle estimator in 2 iterations. The above results can be similarly extended to our DC setting.

1.7 Conclusions

In this work, we close the gap between the statistics and optimization by finding a set of d-stationary solutions to the DC penalized loss functions. Specifically, we relax the assumptions used in [1] and provide stronger statistical results on the penalized estimation problem. We prove that a certain subset of d-stationary solutions in an optimization problem (with a DC objective) has the ideal statistical properties: asymptotic estimation consistency, asymptotic model selection consistency, asymptotic efficiency under linear model and the

GLM settings. We also provide the non-asymptotic upper bound for the estimation errors in both scenarios. We unify the framework of non-convex penalized high-dimensional sparse estimation problems and the existing popular algorithms to solve the problems in a DC framework.

Several open questions remain, which might be interesting directions for future research. Since in this work, we mainly consider the unconstrained DC programming, it is unclear whether a proper constraint, which might depend on specific problems, will ensure a better set of solution or possibly unique solution to the high-dimensional sparse estimation problem. Another direction would be more general loss functions. When the observations have outliers or missing values, it would be desiring to obtain theoretical guarantees on the sparse estimations with possibly non-convex loss functions, such as Huber loss, Cauchy loss, etc.

CHAPTER 2
HIGH-ORDER LAPLACIAN-BASED REGULARIZATION ACHIEVES THE OPTIMAL RATE IN FUNCTION ESTIMATION

2.1 Introduction

Given noisy observations, a standard approach of functional estimation is to introduce a penalty term on the smoothness of the underlying function and make a trade-off between the goodness-of-fit and the penalty. The penalty term typically involves the estimation of the underlying function's derivatives. Among existing literatures, the graph Laplacian related regularization, which converges to the continuous Sobolev semi-norm under specific settings, has been widely used in learning problems to solve the data smoothing problems, which takes advantage of the information on the geometry towards the marginal distribution [43, 44, 45, 46, 47, 48, 49, 50, 51].

In this work, we study the high-order Laplacian regularization of the form: $\mathbf{f}^T\mathbf{L}^m\mathbf{f}$, with $\mathbf{L}$ being the graph Laplacian matrix, where $\mathbf{L}^m$ is the matrix multiplication. We show in Section 2.2.2 that it converges to the continuous Laplacian operator defined semi-norm: $c\int_\Omega f(x)\Delta^m f(x)dx$, under some regularity conditions on the boundary of the domain. It can be easily verified that the high-order Laplacian regularization is an extension of the classic thin-plate splines [52] and soap film smoothers [53]. *Thin-plate Splines* is introduced by [52], which uses the Frobenius norm of the Hessian matrix as the penalty $\int \|\nabla^2 f\|_2^2 dx$, where $\nabla^2 f$ is the Hessian matrix of the function f and $\|\cdot\|_2$ represents the Frobenius norm of a matrix. Wahba (1990) [54] generalizes the Hessian based penalty to a more general Sobolev semi-norm penalty $J_m^d(f) = \sum_{|\alpha|=m}\binom{m}{\alpha}\int_\Omega |D^\alpha f|^2 dx$, which is reviewed in Section 2.2.1, and proves that among nonparametric estimators and in terms of the L_2 risk, the thin-plate estimator achieves the best possible convergence rate that is

given in [2]. More discussion can be found in [48]. On the other hand, the limitation of the thin-plate splines on its capability of handling irregular regions is discussed in [55], which leads to the *soap film smoother* that is proposed in [53]. The regularization term employed in the soap film smoother is $\int (f_{xx} + f_{yy})^2 dxdy$, which is the integral of squared Laplacian instead of the Frobenius norm of Hessian, and therefore admits certain degree of freedom along the boundary of region and therefore it can handle data smoothing over irregular regions. However, the soap film smoother has some drawbacks: *first*, current version of the soap film smoothers lacks theoretical justification in terms of consistency and whether or not it achieves the optimal convergence rate; *second*, soap film smoother involves the solution of PDE's, namely Laplacian equation and Poisson equation, which is not common in machine learning community and it is computationally inefficient in the high dimensional cases. The high-order Laplacian estimator, in comparison, can help alleviate the above disadvantages for both methods with theoretical performance guarantees. It can be viewed as a generalization of the thin plate splines from regular domains to unknown submanifolds, from a coordinate dependent Sobolev semi-norm defined by partial derivatives to a coordinate free high-order Laplacian semi-norm using the Laplacian operators, from fixed data independent reproducing kernels to data dependent kernels [48]. It has the Laplacian-based regularization employed in the thin-plate splines and the soap film smoothers as its special case when the order $m = 2$, with closed form estimation and convergence guarantees, since $\int (f_{xx}+f_{yy})^2 dxdy = \int f \triangle^2 f dxdy$, and $\int \|\nabla^2 f\|_2^2 dx = \int f \triangle^2 f dx$, under some regularity conditions of the domain (seeing Lemma 2.3.1).

High-order Laplacian regularization has been studied in [43, 44, 48]. Its corresponding discrete approximation is based on *graph Laplacian*, which is studied in [56] to capture the local smoothness of the underlying manifold. We review some representative works in the following. Smola and Kondor (2003) [43] propose a family of regularization operators (equivalently, kernels) on graphs that include Diffusion Kernels as a special case, and show that this family encompasses all possible regularization operators invariant under permuta-

tions of the vertices in a particular sense. While the theoretical guarantee of the resulting estimations is missing. Belkin, Matveeva and Niyogi (2004) [44] propose the Tikhonov regularization to label a partially labeled graph by using Laplacian-related regularization. They show that the generalization error is bounded in terms of the smallest nontrivial eigenvalue (Fiedler number) of the graph, by using techniques from algorithmic stability. The theoretical analysis of a Tikhonov regularisation method is conducted regarding to the algorithmic stability. There is no justification on the choice on the regularization parameter. Zhou and Belkin (2011) [48] extend graph Laplacian to high-order Laplacian regularization and utilize this penalty in semi-supervised learning. They show that high-order graph Laplacian approximation converges to its corresponding integral form of high-order Laplacian regularization term, i.e., the consistency of their penalty, and provide some intuition based on theory of reproducing kernel Hilbert space. Still the theoretical justification of high-order Laplacian in terms of convergence rates is not established, in relative to the thin-plate splines in [54]. Zhou, and Srebro (2011), [57] provide connection between integrated mean square error (IMSE) of semi-supervised learning by Laplacian Eigenmaps at the limit of infinite unlabeled data and the graph Laplacian regularizer. They prove that given the exact form of the continuous Laplacian operator, when they take $k = O(n^{-\frac{d}{2+d}})$ eigenfunctions in the Laplacian eigenmaps estimation, they can obtain the asymptotic error rate of $O(n^{-\frac{2}{2+d}})$. There is no guarantee of the asymptotic error rate using the sample data, when the domain of the observation is unknown.

Other related work, which study the graph Laplacian regularization [45, 46, 47, 49, 50, 51], are taking different perspectives from our current work. We discuss them in details at the end of this work. The theoretical justifications of the Laplacian-based regularization in terms of the optimal convergence rate seem to be missing in the literature. In this work, we establish the theoretical justification of high-order Laplacian regularization in terms of convergence rate and the choice of the regularization parameter. Specifically, we consider the data smoothing problem using the least squares loss function and the high-order Laplacian

regularization and name the corresponding estimator *Graph Laplacian Smoother (GLS)*. We consider a general form of transductive learning on graphs with high order Laplacian regularization. We establish the *optimal rate of convergence* for GLS that matches the well-known optimal rate in [2], and therefore provide the first appearance of theoretical justification for the Laplacian-based regularization by considering the operator norms related to the Laplacian operators, which, according to our literature search, has never been used to study similar problems relating to graph Laplacian regularization. Furthermore, we propose the *generalized cross validation* (GCV) to choose the tuning parameter of the regularization, which is not studied in the literature. We establish the asymptotic optimality of GCV under GLS based on the associated Stein's unbiased risk estimates (SURE) [58], which gives a justifiable way of choosing the tuning parameter in the regularization. We define the SURE estimate as in [58]. Along with our analysis on the Laplacian Matrix, we then prove the consistency of GCV.

The rest of the work is organized as follows. In Section 2.2, we review the classical Sobolev semi-norm based regularization [54] and then present our formulation of the graph Laplacian regularization and our computational approach. We study the rate of decaying of the eigenvalues for the corresponding discrete graph Laplacian matrix and prove that the high-order Laplacian regularization reaches the optimal rate of convergence within the nonparametric smoothing in Section 2.3. The asymptotic optimality of GCV is studied in Section 2.3. The main results are established when the marginal distribution of the input variable, X, satisfies a uniform distribution with volume of the domain being equal to 1. However, we extend our results to the cases where X does not necessarily have a uniform distribution with unit volume of the domain. We discuss other directions of graph Laplacian regularization related works in Section 2.4. We conclude this work in Section 2.5. The proofs in Section 2.3 are relegated to the Appendix **??**.

2.2 Methodology

In this section, we formulate the problem of *high-order Laplacian* regularization in the context of supervised learning. We review the general framework adopted in the smoothing splines and the classic Sobolev semi-norm regularization [54] in Section 2.2.1. Then, we present the high-order Laplacian regularization and the corresponding computational scheme in details in Section 2.2.2. The *generalized cross validation* (GCV) approach for choosing the optimal penalty parameter λ is described in Section 2.2.3.

2.2.1 General Framework

Let $(X_i, y_i), i = 1, 2, \cdots, n$ denote the observations. We assume that the data generation mechanism is

$$y_i = f(X_i) + \varepsilon_i,$$

where y_i's are the responses, $X_i \in \mathbb{R}^d$, $i = 1, 2, \cdots, n$, are the predictors, and ε_i's are independent noises with mean 0 and variance σ^2. The capitalized X_i indicates that it can be multivariate. Function estimation is to uncover $f(\cdot)$, within a known domain $\Omega \subset \mathbb{R}^d$. To search for a function $\hat{f}$ in a function space $\mathcal{F}$ such that it is a reasonable estimate of the true underlying function f, a standard approach is to minimize a functional as follows:

$$\min_{f \in \mathcal{F}} \quad \sum_{i=1}^{n} \big(y_i - f(X_i)\big)^2 + \lambda \mathcal{J}(f), \tag{2.2.1}$$

where the first term is called a goodness-of-fit measure and the second term penalizes the smoothness of the estimation. The above optimization can be considered as the trade-off between the estimation error and the model complexity. Difference between many smoothers lies in the choice of the regularization term, namely $\mathcal{J}(f)$. Below, we review a classic regularizer defined by a Sobolev semi-norm (seeing [54]).

We consider the functional space $\mathcal{F}$ to be a Sobolev space on Ω. More specifically,

let $\mathbb{Z}_+^d$ denote the set of all ordered d-tuples of nonnegative integers. For $\alpha \in \mathbb{Z}_+^d$, $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_d)$ and $|\alpha| = \sum_{i=1}^d \alpha_i$, the partial derivative is defined as

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}}.$$

Let $\Omega \subset \mathbb{R}^d$ denote the domain of the functions. Then the Sobolev space of order m, denoted by $W^{m,2}(\Omega)(= H^m(\Omega))$, is defined to be the space consisting of those functions in $L^2(\Omega)$ that, together with all their weak partial derivatives up to and including those of order m, belong to $L^2(\Omega)$, i.e., we have

$$H^m(\Omega) = \left\{ f : D^\alpha f \in L^2(\Omega), \forall \alpha \in \mathbb{Z}_+^d, |\alpha| \leq m \right\}.$$

Define Sobolev semi-inner product as in [54]:

$$\langle u, v \rangle_m = \sum_{|\alpha|=m} \binom{m}{\alpha} \int_\Omega D^\alpha u D^\alpha v dx, \tag{2.2.2}$$

where $\binom{m}{\alpha} = \frac{m!}{\alpha_1! \cdots \alpha_d!}$. The induced Sobolev semi-norm [54] is

$$J_m^d(f) = \sum_{|\alpha|=m} \binom{m}{\alpha} \int_\Omega |D^\alpha f|^2 dx = \sum_{|\alpha|=m} \binom{m}{\alpha} \|D^\alpha f\|_{L^2}^2, \tag{2.2.3}$$

and $J_m^d(\cdot)$ is shown to be a good choice in the sense that the estimated function f is continuous and lies in the corresponding RKHS (Reproducing Kernel Hilbert Space) if $2m > d$. It has nice theoretical foundation rooted in functional analysis and the theory of Sobolev space. It is equivalent to the high-order Laplacian regularization in the limiting continuous scenario [48] with proper boundary conditions, which we will discuss in Section 2.2.2.

2.2.2 Problem Formulation

In this work, the following high-order Laplacian regularization is studied,

$$\min_{f\in\mathcal{F}} \quad \frac{1}{n}\sum_{i=1}^{n}\big(y_i - f(X_i)\big)^2 + \lambda I_m^d(f), \tag{2.2.4}$$

where $I_m^d(f) = |f|_{\Omega,m}^2 = \int_\Omega f(x)\Delta^m f(x)dx$, $\Delta = -\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}$ and $\Delta^m f(x) = \Delta(\Delta^{m-1}f(x))$ representing applying the Laplacian operator (i.e., Δ) m times to the function $f(x)$.

Discrete approximation under uniform distribution

When the density of $X \in \Omega$ is a uniform distribution, with $\mathrm{Vol}(\Omega) = V$, the following discrete approximation can be easily verified thanks to the theory in [56].

Theorem 2.2.1. *Let Ω be a compact connected submanifold in $\mathbb{R}^d$ without boundary. $x_1, \cdots, x_n$ are sampled uniformly on Ω, and m be a positive integer. Assume $f \in C^{2m}(\Omega), \mathrm{Vol}(\Omega) = V$, then for $t_n = O(n^{-\frac{1}{d+2+\alpha}})$ where $\alpha > 0$ as $n \to \infty$ we have*

$$\frac{c}{n}\left(\frac{1}{nVt_n^{d/2+1}}\right)^m \mathbf{f}^T\mathbf{L}^m\mathbf{f} \xrightarrow{p} \int_\Omega f(x)\Delta^m f(x)dx = I_m^d(f) \tag{2.2.5}$$

where t_n is the bandwidth of a Gaussian kernel function and c is a constant.

Specifically, when the domain Ω has a unit volume, we have the following special result, which also appears in Theorem 4 from [48]:

Theorem 2.2.2. *Let Ω be a compact connected submanifold in $\mathbb{R}^d$ without boundary. $x_1, \cdots, x_n$ are sampled uniformly on Ω, and m be a positive integer. Assume $f \in C^{2m}(\Omega), \mathrm{Vol}(\Omega) = 1$, then for $t_n = O(n^{-\frac{1}{d+2+\alpha}})$ where $\alpha > 0$ as $n \to \infty$ we have*

$$\frac{c}{n}\left(\frac{1}{nt_n^{d/2+1}}\right)^m \mathbf{f}^T\mathbf{L}^m\mathbf{f} \xrightarrow{p} \int_\Omega f(x)\Delta^m f(x)dx = I_m^d(f) \tag{2.2.6}$$

where t_n is the bandwidth of a Gaussian kernel function and c is a constant.

Discrete approximation under non-uniform distribution

When the density on Ω is not uniform, the limit to the convergence of the graph Laplacian can be obtained similarly [56]. However, it uses the standardized weights in the discrete Laplacian matrix, and the limit is a weighted Laplacian with regard to the probability density. Let $\mathbf{W}^s$ represent the standardized weight matrix, where the (i,j)th entry is given by

$$w^s_{i,j} = \frac{1}{t}\frac{G_t(x_i,x_j)}{\sqrt{\hat{d}_t(x_i)}\sqrt{\hat{d}_t(x_j)}},$$

where $G_t(x_i,x_j) = \frac{1}{(4\pi t)^{d/2}} e^{-\frac{\|x_i-x_j\|^2}{4t}}$ is the Gaussian weight, and $\hat{d}_t(x_i) = \frac{1}{n-1}\sum_{j\neq i} G_t(x_i,x_j)$. Let L^t_n denote the corresponding Laplacian operator

$$L^t_n f(x) = \frac{1}{n}\sum_{i=1}^{n} w^s(x,x_i)(f(x)-f(x_i)).$$

We will have the following limit of $L^t_n f(x)$ according to [56]:

$$L^t_n f(x) \xrightarrow{p} \Delta_P f(x), \tag{2.2.7}$$

where $P(x)$ is the probability density function in Ω and
$\Delta_P f(x) = \frac{1}{P(x)}\text{div}(P(x)\nabla f(x))$ is the weighted Laplacian corresponding to $P(x)$. More details on the proof of the above can be found in [56].

According to the results in Subsection 2.2.2 and 2.2.2, for simplicity in our statement and proofs, we will state all our results in the case of uniform distribution on a unit volume domain. Thanks to the theory in [56], the discrete approximation to the term $I^d_m(f)$ (which is proposed in [48]) can be as follows:

$$I^d_{m,n}(f) = \frac{1}{n}\left(\frac{1}{nt_n^{d/2+1}}\right)^m \mathbf{f}^T\mathbf{L}^m\mathbf{f}, \tag{2.2.8}$$

where $\mathbf{f} = (f(X_1),\cdots,f(X_n))^T$, $\mathbf{L} = \mathbf{D}-\mathbf{W}$ is the graph Laplacian matrix (seeing [59]),

$\mathbf{D}$ is a diagonal matrix with diagonal element $D_{i,i} = \sum_j w_{i,j}$ and $\mathbf{W} = (w_{i,j})_{1\le i,j\le n}$, where $w_{i,j} = e^{-\frac{\|x_i - x_j\|^2}{t}}$, t is the bandwidth that depends on n. The structure of $w_{i,j}$ is essentially the Gaussian kernel which imposes the local relation that depends on the bandwidth t.

Throughout this section and Section 2.3, we assume that the sample data $T = \{X_i\}_{i=1}^n$ are uniformly drawn from Ω, i.e., the marginal distribution of X is the uniform distribution on Ω.

Therefore, our discrete approximation to (2.2.4) is as follows

$$\min_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda I^d_{m,n}(f), \tag{2.2.9}$$

which is equivalent to

$$\min_{f\in\mathcal{F}} (\mathbf{y}-\mathbf{f})^{\mathbf{T}}(\mathbf{y}-\mathbf{f}) + \frac{\lambda}{n^m t^{m(d/2+1)}}\mathbf{f}^{\mathbf{T}}\mathbf{L}^{\mathbf{m}}\mathbf{f}. \tag{2.2.10}$$

Let $\mathbf{M} = (nt^{d/2+1})^{-m}\mathbf{L}^m$. In order to study the optimality of our estimator

$$\hat{\mathbf{f}} = (I + \lambda\mathbf{M})^{-1}\mathbf{y}, \tag{2.2.11}$$

which is derived from the first order condition of optimizing (2.2.10), it is equivalent to study the properties of eigenvalues of matrix $\mathbf{M}$, which will be fully examined in Section 2.3.

Given Theorem 2.2.1, in order for the results on the convergence rate and GCV in Section 2.3 to still hold, we only need some minor modifications to Theorem 2.3.3 to bound eigenvalues of the new matrix

$$\mathbf{M}' = (nVt^{d/2+1})^{-m}\mathbf{L}^m.$$

Results in Section 2.3.1 are all about the properties with regard to the newly defined semi-

norms, which only involves the use of continuous Laplacian operator. They still hold in this new scenario.

For the non-uniform distribution scenario, the key is still to bound the eigenvalues of the corresponding matrix, derived from

$$L_n^t f(x) = \frac{1}{n}\sum_{i=1}^{n} w^s(x, x_i)(f(x) - f(x_i)).$$

While in order to prove the bound on the eigenvalues, we will need to study an Laplacian induced semi-norm. We will give theorems on this in Section 2.3 and prove that all the results still hold in this case.

2.2.3 Choice of the Penalty Parameter λ

Besides studying the convergence on the estimation, we further adopt the generalized cross validation (GCV) for the linear estimation to determine the optimal penalty parameter λ. Specifically, the generalized cross validation function is defined as

$$\mathrm{GCV}_n(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{f}_{n,\lambda}(X_i)}{1 - \frac{1}{n}\mathrm{tr}[\mathbf{A_n}(\lambda)]}\right)^2,$$

where $\mathbf{A}_n(\lambda) = (\mathbf{I}_n + \lambda\mathbf{M})^{-1}$. The optimal value of the penalty parameter λ can be estimated by minimizing the above GCV function, i.e.,

$$\hat{\lambda}_G = \mathrm{argmin}_{\lambda>0}\mathrm{GCV}_n(\lambda).$$

For practical purpose, we could re-parameterize the $\lambda = e^{\theta}$, $\theta \in \mathbb{R}$ to convert the minimization into a unconstrained optimization problem. The justification is relatively straightforward and can be found in [60].

2.3 Theoretical Properties

In this section, we establish the optimal convergence rate for the GLS (graph Laplacian smoother) that is introduced in Section 2.2 and the asymptotic optimality of the GCV. Section 2.3.1 provides the necessary mathematical foundations, mainly, some inequalities from functional analysis, as the preparation to study the optimal convergence rate. In Section 2.3.2, we discuss the asymptotic rates of the matrix $\mathbf{M}$'s eigenvalues using the tools developed in Section 2.3.1. We derive its asymptotic properties from different aspects. The multivariate input situation ($d > 1$) is included in Section 2.3.3, where we show that asymptotic properties comparable to smoothing splines in [54]. Finally we introduce the asymptotic optimality of the GCV, and show in Section 2.3.4 that for GLS, the asymptotic optimality of GCV is preserved, hence GCV is a justifiable way for choosing the parameter λ.

2.3.1 Mathematical Preparation

From Section 2.2, we pin down our problem to bound the eigenvalues of the penalty matrix $\mathbf{M}$. Since $\mathbf{M} \propto \mathbf{L}^m$ and L is a positive semi-definite symmetric matrix, we will first consider the bounds of eigenvalues of $\mathbf{M}$ defined in Section 2.2.2 when $m = 1$. We then extend it to the case of $m > 1$.

For any domain $\Omega \subset \mathbb{R}^d$, which satisfies the conditions in Lemma 2.3.1 below, let $H^m(\Omega)$ denote the mth-order Sobolev space of the generalized functions. As in [48], we define the semi-inner product by

$$\langle f, g\rangle_{\Omega,m} = \int_\Omega f(x)\Delta^m g(x)dx = \int_\Omega g(x)\Delta^m f(x)dx, \tag{2.3.1}$$

which gives rise to the related semi-norm

$$|f|^2_{\Omega,m} = \int_\Omega f(x)\Delta^m f(x)dx \tag{2.3.2}$$

The properness of the semi-norm is ensured by the following lemma, which also appears in [48].

Lemma 2.3.1. *$|f|_{\Omega,m}$ is a semi-norm given one of the following two conditions:*

1. $\partial\Omega = \emptyset$,

2. $\nabla(\Delta^k f(x)) \cdot \boldsymbol{n} = 0, \forall x \in \partial\Omega, k = 0, 1, \cdots, m-1$,

where $\partial\Omega$ denotes the boundary of Ω and ∇ stands for the gradient operator and $\mathbf{n}$ is the normal vector orthogonal to $\partial\Omega$.

The proof is a direct use of the Green's identity and is postponed to Appendix B.1. We list the proof, because a proof is not included in the aforementioned original reference. The first condition requires the empty boundary of Ω, which corresponds to the *closed submanifold* in the Euclidean space. In this work, we focus on the second case. This requirement of boundary is quite similar to the *Neumann Boundary Condition* (also appeared in [55]), which has a nice physical meaning (see Appendix B.22 for further illustration.).

In order to prove our main results with regard to weighted Laplacian operator in Section 2.2.2, the properties in Section 2.3.1 need major modifications towards $\Delta_P(\cdot)$. According to [61, 56], we have the following results on the weighted Laplacian operator.

Theorem 2.3.1. *Let $\mathcal{M}$ represent a manifold with measure $d\nu$. For any probability density function $P(x)$ defined on $\mathcal{M}$, let $d\mu = Pd\nu$, we have*

$$\int_{\mathcal{M}} u\Delta_P v d\mu = \int_{\mathcal{M}} \langle \nabla u, \nabla v \rangle d\mu.$$

From Theorem 2.3.1, we can see the weighted Laplacian operator is semi-positive definite. Thus, the weighted Laplacian operator defined on a weighted manifold enjoys the same properties for Laplacian operator defined the manifold. All the results in Section 2.3.1 can be similarly proved under the weighted scenario. We will have the same convergence and GCV results as proved for the uniform distribution case.

With observations $T = \{X_i\}_{i=1}^n$, we can define a discrete version of the aforementioned semi-norm as

$$|f|_{T,m}^2 = \frac{1}{n}\sum_{i=1}^{n} f(X_i)\Delta^m f(X_i). \tag{2.3.3}$$

When $m = 0$, $|f|_{T,0}^2 = \frac{1}{n}\sum_{i=1}^n f(X_i)^2$ and $|f|_{\Omega,0}^2 = \int_\Omega f^2(x)dx$. Let $\mathbf{E}_{T,1}$ be the representing matrix such that

$$|f|_{T,1}^2 = \frac{1}{n}\mathbf{f}^T\mathbf{E}_{T,1}\mathbf{f},$$

where $\mathbf{f} = (f(X_1), \cdots, f(X_n))^T$ and f corresponds to the solution of a variational problem:

$$|f|_{T,1} = \min_{\phi\in H^1(\Omega),\phi(X_i)=y_i} |\phi|_{T,1} \tag{2.3.4}$$

The existence of matrix $\mathbf{E}_{T,1}$ is established as follows. Please refer to Appendix B.2 for a detailed proof.

Proposition 2.3.1. *There exists a matrix $\mathbf{E}_{T,m}$, such that*

$$|f|_{T,m}^2 = \min_{\phi\in H^m(\Omega),\phi(X_i)=y_i} |\phi|_{T,m}^2 = \frac{1}{n}y^T\mathbf{E}_{T,m}y \tag{2.3.5}$$

where $T = \{X_i\}_{i=1}^n$, $\mathbf{y} = (y_1, y_2, .., y_n)^T = (f(X_1), f(X_2), \cdots, f(X_n))^T$.

In our main results (Theorem 2.3.5), where we provide the optimal rate of convergence rate for the estimator defined (2.2.11), we mainly rely on the bound for the eigenvalues of matrix $\mathbf{M}$. This can be obtained from the bound of the eigenvalues of the graph Laplacian matrix. We will prove in Section 2.3.2 that the eigenvalues of $\mathbf{M}$ and that of $\mathbf{E}_{T,m}$ are equal as the sample size goes to infinity, which further assists the proof of our main results. In the following Lemma 2.3.2, we prove the properties for $\mathbf{E}_{T,1}$. For the set of sampling points $T = \{X_i\}_{i=1}^n$ in domain Ω, we assume that there exists a constant $B_0 > 0$ such that $\delta_{\max}/\delta_{\min} \leq B_0$, where $\delta_{\max} = \sup_{X\in\Omega}\inf_{X_i\in T}\|X - X_i\|$, and $\delta_{\min} = \min_{j\neq i}\|X_j - X_i\|$. Next we establish some properties needed for domain Ω. The proof is included in Appendix B.3.

Lemma 2.3.2. *If Ω is a bounded domain in $\mathbb{R}^d$ and satisfies the condition in Lemma 2.3.1 and $\delta_{\max}/\delta_{\min} \leq B_0$. Denote e_n as the largest eigenvalue of matrix $\boldsymbol{E}_{T,1}$, then $n\delta_{\max}^d$ and $\delta_{\max}^2 e_n$ are both bounded from above, i.e., there exist constants B_1, B_2 such that $n\delta_{\max}^d \leq B_1$ and $\delta_{\max}^2 e_n \leq B_2$.*

Recall that a domain with Lipschitz boundary is a set in Euclidean space whose boundary is sufficiently regular in the sense that it can be thought of as locally being the graph of a Lipschitz continuous function. Let Ω be an open set in $\mathbb{R}^d$ satisfying a *uniform cone condition*, which is defined as follows.

Definition 2.3.1. *An open set Ω in $\mathbb{R}^d$ is said to satisfy the uniform cone condition, if there exists a radius $r > 0$ and an angle $\theta \in (0, \pi/2)$ such that for any $X \in \Omega$, a unit vector $\zeta(X) \in \mathbb{R}^d$ exists such that the cone*

$$C(X, \zeta(X), r, \theta) = \{X + t\mathbf{s} : \mathbf{s} \in \mathbb{R}^d, \|\mathbf{s}\| = 1, \zeta(X)^T\mathbf{s} \geq \cos\theta, 0 \leq t \leq r\}$$

is contained in Ω.

To exhibit the Rayleigh quotient inequalities connecting Sobolev semi-norms (same as our definition when $m = 1$) and their discretized version, let

$$U_2^1(\Omega) = \{f \in H^1(\Omega) : \underline{B}|f|_{\Omega,1}^2 \leq |f|_{T,1}^2 \leq \bar{B}|f|_{\Omega,1}^2\}$$

denote a class of functions with bilaterally bounded constraint on their first order derivative, where $\underline{B}, \overline{B}$ are independent of f. In the univariate case, we prove that the function class $U_2^1(\Omega)$ contains the polynomial spline space of degree $m+1$ with knots at $T = \{X_i\}_{i=1}^n$. We put this auxiliary results in Appendix B.4. Notice that

$$|f|_{\Omega,1}^2 = \int_\Omega f(x)\Delta f(x)dx = \int_\Omega \|\nabla f(x)\|^2 dx = \int_\Omega \sum_{|\alpha|=1} |D^\alpha f(x)|^2 dx$$

under conditions in Lemma 2.3.1, and this is the same as the classic definition of semi-norm in Sobolev space. In order to examine the connection between the continuous and discrete version of semi-norms, we have the following lemmas, which are key steps to establish the closeness of eigenvalues of $\mathbf{E}_{T,1}$ and spectrals of elliptic operator Δ.

Lemma 2.3.3. *Let Ω be an open bounded Lipschitz domain satisfying both uniform cone condition and conditions in Lemma 2.3.1. Then there exists constant $C_1 = C_1(d, \Omega, B_0, \underline{B}) > 0$ and $\delta_0 > 0$ such that if $\delta_{\max} \leq \delta_0$, we have the following for any $f \in U_2^1(\Omega)$,*

$$\frac{|f|_{T,1}^2}{|f|_{T,0}^2} \geq \frac{|f|_{\Omega,1}^2}{C_1(|f|_{\Omega,0}^2 + \delta_{\max}^2 |f|_{\Omega,1}^2)} \tag{2.3.6}$$

Lemma 2.3.4. *Let Ω be an open bounded Lipschitz domain satisfying both uniform cone condition and conditions in Lemma 2.3.1. Then there exists constant $C_2 = C_2(d, \Omega, B_0, \underline{B}, \overline{B}) > 0$ and $\delta_0 > 0$ such that if $\delta_{\max} \leq \delta_0$, we have the following for any $f \in U_2^1(\Omega)$,*

$$\frac{|f|_{\Omega,1}^2}{|f|_{\Omega,0}^2} \geq \frac{|f|_{T,1}^2}{C_2(|f|_{T,0}^2 + \delta_{\max}^2 |f|_{T,1}^2)} \tag{2.3.7}$$

Let $e_1 \leq \cdots \leq e_n$ be the eigenvalues of $\mathbf{E}_{T,1}$ in an ascending order. Clearly $\{e_j\}_{j=1}^n$ are non-negative real numbers since the matrix $\mathbf{E}_{T,1}$ is semi-positive definite. Next we will establish the convergence rate of the eigenvalues and show that they can be bounded by the discrete spectrum of the first order Laplacian Δ^1.

Lemma 2.3.5. *Let Ω be an open bounded Lipschitz domain satisfying both uniform cone condition and conditions in Lemma 2.3.1. Then there exists constants $C_3, C_4 > 0$ such that*

$$C_3 \rho_j \leq e_j \leq C_4 \rho_j$$

where $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_n$ are the first n eigenvalues of the variational eigenvalue problem

$$\langle \phi, \psi \rangle_{\Omega,1} = \rho \langle \phi, \psi \rangle_{\Omega,0} \quad \forall \psi \in H_2^1(\Omega), \phi \in U_2^1(\Omega)$$

The proofs of Lemma 2.3.3, 2.3.4, 2.3.5 can be found in Appendix B.5, B.6, B.7. Based on the above lemmas, we have one of our main results:

Theorem 2.3.2. *Let Ω be an open bounded Lipschitz domain satisfying both uniform cone condition and conditions in Lemma 2.3.1. Recall $e_1 \leq e_2 \leq \cdots \leq e_n$ are the eigenvalues of $\boldsymbol{E}_{T,1}$ in ascending order. Then there exists constants $C_5, C_6 > 0$ such that for $1 < j \leq n$, we have*

$$C_5 j^{2/d} \leq e_j \leq C_6 j^{2/d}$$

Proof. Apply Theorem 14.6 in [62] to get $\rho_j \sim j^{2/d}$ if $j > 1$. And with result from Lemma 2.3.5, it concludes the proof. □

We include Theorem 14.6 of [62] in Appendix B.21 for completeness. Here we require $j > 1$ because for the Laplacian operator Δ, its smallest eigenvalue $\rho_1 = 0$ and the corresponding eigenfunction ϕ_1 spans the null space of this operator. Correspondingly, it is obvious that $\mathbf{L}$ also has 0 as its eigenvalue and vector $\mathbf{1}$ as the corresponding eigenvector.

2.3.2 Bounds of regularization matrix M's eigenvalues

In this section, we use the previous results on the eigenvalues of $\mathbf{E}_{T,1}$ to study the eigenvalues of the graph Laplacian matrix $\mathbf{L}$. The basic idea is to build a connection between matrix $\mathbf{L}$ and $\mathbf{E}_{T,1}$ (w.r.t. $m = 1$). Then the bounds of $\mathbf{M}(\propto \mathbf{L}^m)$'s eigenvalues can be easily obtained.

In order to achieve this, we first list some properties of matrix $\mathbf{L}$:

1. Symmetricity, $\mathbf{L} = \mathbf{L}^T$.

2. $\mathbf{L} \succeq 0$ and the smallest eigenvalue equals 0. It is the discrete approximation of the continuous Laplacian operator Δ.

3. $\mathbf{L} = \frac{1}{2}\sum_{i=1}^{n} \mathbf{L}_i$, where

$$\mathbf{L}_i = \begin{pmatrix} w_{i,1} & 0 & 0 & -w_{i,1} & 0 & 0 & 0 \\ 0 & \ddots & 0 & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & w_{i,i-1} & -w_{i,i-1} & 0 & 0 & 0 \\ -w_{i,1} & \cdots & -w_{i,i-1} & \sum_{j:j\neq i} w_{i,j} & -w_{i,i+1} & \cdots & -w_{i,n} \\ 0 & \cdots & 0 & -w_{i,i+1} & w_{i,i+1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & -w_{i,n} & 0 & \cdots & w_{i,n} \end{pmatrix}$$

The mathematical intuition of defining $\mathbf{L}_i$ is that it uses all sampled data points with kernel function to approximate the Laplacian around X_i. It becomes more clear if we consider

$$\mathbf{L}_i\mathbf{f} = (w_{i,1}f_1 - w_{i,1}f_i, \cdots, \sum_{j:j\neq i} w_{i,j}f_i - \sum_{j:j\neq i} w_{i,j}f_j, \cdots, w_{i,n}f_n - w_{i,n}f_i)^T$$

and the following lemmas. The proofs are quite similar to those from [56], which considers more general form on manifolds. Here we consider the case in the Euclidean space.

Lemma 2.3.6. *Given any open ball $B \subset \Omega$ and $p \in B$, for any $l \in \mathbb{N}$, it holds that as $t \to 0$,*

$$\int_B e^{-\frac{\|p-y\|^2}{4t}} f(y)dy - \int_\Omega e^{-\frac{\|p-y\|^2}{4t}} f(y)dy = o(t^l) \tag{2.3.8}$$

Lemma 2.3.6 proves the fact that we can use only a small open set (i.e., the Euclidean ball in our case) around a fixed data point to estimate the heat kernel over Ω at that point, and the error decays exponentially as the bandwidth shrinks. This will be useful when we establish the local approximation of the Laplacian operator.

Lemma 2.3.7. *There exists a constant C such that*

$$\frac{\partial}{\partial t}\left((4\pi t)^{-\frac{d}{2}}\int_{B(p)} e^{-\frac{\|p-y\|^2}{4t}} f(y)dy\right)\Bigg|_{t=0} = \Delta f(p) + Cf(p) \tag{2.3.9}$$

Lemma 2.3.7 establishes the connection between the heat kernel and the Laplacian operator. Its proof is the same as that of Lemma 9 in [56]. The combination of Law of Large Numbers (LLN) and the following Lemma 2.3.8 shows that the Laplacian at each data point can be approximated by weighted average of function values at sampled data points, which will be used to establish the asymptotic properties of the matrix $\mathbf{M}$.

Lemma 2.3.8.

$$\lim_{t\to 0}(\pi t)^{-d/2-1}\left(\int_\Omega w_{i,j} f(x_i)dx_j - \int_\Omega w_{i,j} f(x_j)dx_j\right) = \Delta f(x_i)$$

where $w_{i,j} = e^{-\frac{\|x_i - x_j\|^2}{t}}$ is the value of gaussian kernel between x_i and x_j.

Now we can state the main result regarding the eigenvalues of matrix $\mathbf{M}$:

Theorem 2.3.3. *Let $\mu_1 \leq \cdots \leq \mu_n$ be the eigenvalues of the matrix $\mathbf{M}$. There exists constants $C_7, C_8 > 0$ such that for $1 < j \leq n$ we have*

$$C_7 j^{2m/d} \leq \mu_j \leq C_8 j^{2m/d}.$$

For the completeness of the work, we now re-state the results to bound eigenvalues of the matrix $\mathbf{M}'$, which is defined in Section 2.2.2 for the uniform non-unit domain, as follows.

Theorem 2.3.4. *Let $\mu_1 \leq \cdots \leq \mu_n$ be the eigenvalues of the matrix $\mathbf{M}'$. There exists constants $C_9, C_{10} > 0$ such that for $1 < j \leq n$ we have*

$$C_9 j^{2m/d} \leq \mu_j \leq C_{10} j^{2m/d}.$$

The proof is similar to the proof for Theorem 2.3.3 and is relegated to the Appendix B.20. All other results in Section 2.3.3 and 2.3.4 still hold with the matrix

$$\mathbf{A}_n'(\lambda)\mathbf{y} = (\mathbf{I_n} + \lambda\mathbf{M}')^{-\mathbf{1}}.$$

In summary, all the results on the convergence rate and GCV still hold for uniform distributions with non-unit volumes.

2.3.3 Convergence Rate of Multivariate GLS Estimator

In this section, we prove the optimal rate of convergence for GLS estimation, based on the asymptotic properties of the matrix $\mathbf{M}$'s eigenvalues. The following theorem is the main result in this section:

Theorem 2.3.5. *Let* $\hat{\mathbf{f}}_n(\lambda) = \boldsymbol{A}_n(\lambda)\mathbf{y} = (\mathbf{I_n} + \lambda\mathbf{M})^{-\mathbf{1}}\mathbf{y}$ *be the estimator of the Laplacian regularizer with the order* $m > d/2$ *and denote* $r_n(\lambda) = n^{-1}\|\hat{\mathbf{f}}_n(\lambda) - \mathbf{f}\|^{\mathbf{2}}$. *If* $n \to \infty$ *and* $\lambda \sim n^{-2m/(2m+d)}$ *is chosen, then*

$$\mathbb{E}[r_n(\lambda)] = O(n^{-\frac{2m}{2m+d}}) \tag{2.3.10}$$

In particular, when the function $f \in H^m(\Omega)$, if the smoothing parameter is chosen to satisfy $\lambda \sim n^{-2m/(2m+d)}$, we achieve the convergence rate $\mathbb{E}[r_n(\lambda)] = O(n^{-\frac{2m}{2m+d}})$, which is the optimal convergence rate for multivariate function estimation with the order m in a d-dimensional space ([2]).

2.3.4 Asymptotic Optimality of GCV

In this section, we show that the proposed GLS satisfies some general conditions and then prove the asymptotic optimality of GCV under our proposed framework.

General Conditions

Let $\hat{\mathbf{f}}_n(\lambda) = \mathbf{A_n}(\lambda)\mathbf{y} = (\mathbf{I_n} + \lambda\mathbf{M})^{-\mathbf{1}}\mathbf{y}$ be the estimator from smoothing splines of high-order Laplacian regularization with order m and denote $r_n(\lambda) = n^{-1}\|\hat{\mathbf{f}}_n(\lambda) - \mathbf{f}\|^{\mathbf{2}}$. The asymptotic optimality of GCV is defined as

$$\frac{r_n(\hat{\lambda}_G)}{\inf_{\lambda\in\mathbb{R}_+} r_n(\lambda)} \xrightarrow{p} 1 \tag{2.3.11}$$

which verifies the closeness between the values of risk function given by the GCV choice $\hat{\lambda}_G$ and theoretically optimal choice λ^*, where $\lambda^* = \arg\inf_{\lambda\in\mathbb{R}_+} r_n(\lambda)$.

The main result of this section is to show that the GLS satisfies the following three conditions.

(A.1) $\inf_{\lambda\in\mathbb{R}_+} \mathbb{E}[nr_n(\lambda)] \to \infty$.

(A.2) There exists a sequence $\{\lambda_n\}$ such that $r_n(\lambda_n) \xrightarrow{p} 0$.

(A.3) Let $0 \le \kappa_1 \le \cdots \le \kappa_n$ be the eigenvalues of $\mathbf{K_n}(\lambda) = \lambda\mathbf{M}$. For any l such that $l/n \to 0$, then as $n \to \infty$,

$$\frac{(n^{-1}\sum_{i=l+1}^{n} \kappa_i^{-1})^2}{n^{-1}\sum_{i=l+1}^{n} \kappa_i^2} \to 0$$

From Theorem 2.3.3, we already know that $\mu_1 = 0$ and denote the null space spanned by the first eigenfunction of Δ as $\mathcal{N}$. Then we have the following, which provides the verification of condition (A.1).

Lemma 2.3.9. *If $f \notin \mathcal{N}$, the estimator $\hat{\mathbf{f}}_n(\lambda)$ has the property:*

$$\inf_{\lambda\in\mathbb{R}_+} \mathbb{E}[nr_n(\lambda)] \to \infty,$$

which is (A.1).

In order to establish the bounds on errors, it would be convenient if the random term $r_n(\lambda) = n^{-1}\|\hat{\mathbf{f}}_n(\lambda) - \mathbf{f}\|^2$ can be replaced by its expectation, which is deterministic. For-

tunately, it is true and we state it in the following Lemma:

Lemma 2.3.10. *Under condition (A.1), we have*

$$\sup_{\lambda>0}\left|\frac{r_n(\lambda)}{E[r_n(\lambda)]}-1\right|\to 0.$$

The condition (A.2) shows that the risk function $r_n(\lambda_n)$ converges to zero in probability with proper sequence $\{\lambda_n\}$. From Theorem 2.3.5, we know $\mathbb{E}[r_n(\lambda)]\to 0$ as $n\to\infty$, if $\lambda\sim n^{-\frac{2m}{2m+d}}$. Besides, $r_n(\lambda)$ and its expectation are "close" according to Lemma 2.3.10, therefore (A.2) holds true.

Again since $\mu_j = O(j^{2m/d})$, we have the following lemma.

Lemma 2.3.11. *In our model, for any l such that $l/n\to 0$ and $\kappa_{l+1}>0$, the ratio*

$$\frac{(n^{-1}\sum_{i=l+1}^{n}\kappa_i^{-1})^2}{n^{-1}\sum_{i=l+1}^{n}\kappa_i^{2}}\to 0, \qquad \textit{as} \quad n\to\infty.$$

This verifies the condition (A.3) and is an intermediate result used in Section 2.3.4 and it plays an important role in the asymptotic analysis. The proofs of Lemma 2.3.9, 2.3.10, and 2.3.11 are provided in Appendix B.12, B.13, and B.14, respectively.

Asymptotic Optimality Theorem

Under the aforementioned three conditions (i.e., (A.1)-(A.3)), we will prove the asymptotic optimality of GCV.

Lemma 2.3.12. *Under the condition (A.2), we have*

$$n^{-1}\mathrm{tr}[\mathbf{I}_n-\mathbf{A}_n(\lambda_n)]\to 1, \tag{2.3.12}$$

and

$$n^{-1}\|(\mathbf{I}_n-\mathbf{A}_n(\lambda_n))\mathbf{y}\|^2\to\sigma^2. \tag{2.3.13}$$

The asymptotic results in Lemma 2.3.12 will be used in the proof of Lemma 2.3.14.

Lemma 2.3.13. *Under the condition (A.3), for a sequence λ_n such that $r_n(\lambda_n) \to 0$, we have*

$$\frac{(n^{-1}\mathrm{tr}[\mathbf{A_n}(\lambda_\mathbf{n})])^\mathbf{2}}{n^{-1}\mathrm{tr}[\mathbf{A_n}(\lambda_\mathbf{n})^\mathbf{2}]} \to 0. \tag{2.3.14}$$

Finally we build the connection between $r_n(\lambda)$ and $\mathrm{SURE}_n(\lambda)$, and its proof can be found in Appendix B.17.

Lemma 2.3.14. *For any $\hat{\lambda}$ such that $r_n(\lambda) \to 0$ and*

$$\frac{(n^{-1}\mathrm{tr}[\mathbf{A_n}(\hat{\lambda})])^\mathbf{2}}{n^{-1}\mathrm{tr}[\mathbf{A_n}(\hat{\lambda})^\mathbf{2}]} \to 0 \tag{2.3.15}$$

under the condition (A.1), we have

$$\frac{\left|\mathrm{SURE}_n(\hat{\lambda}) - \tilde{r}_n(\hat{\lambda}) - n^{-1}\|\varepsilon\|^2 + \sigma^2\right|}{r_n(\hat{\lambda})} \xrightarrow{p} 0, \tag{2.3.16}$$

and

$$\frac{n^{-1}\|\tilde{\mathbf{f}}_n(\hat{\lambda}) - \hat{\mathbf{f}}_n(\hat{\lambda})\|^2}{r_n(\hat{\lambda})} \xrightarrow{p} 0, \tag{2.3.17}$$

where

$$\begin{aligned}
\mathrm{SURE}_n(\lambda) &= \sigma^2 - \sigma^4 \frac{(n^{-1}\mathrm{tr}[\mathbf{I_n} - \mathbf{A_n}(\lambda)])^\mathbf{2}}{n^{-1}\|(\mathbf{I_n} - \mathbf{A_n}(\lambda))\mathbf{y}\|^\mathbf{2}}, \\
\tilde{\mathbf{f}}_n(\lambda) &= \mathbf{y} - \sigma^\mathbf{2} \frac{\mathrm{tr}[\mathbf{I_n} - \mathbf{A_n}(\lambda)]}{\|(\mathbf{I_n} - \mathbf{A_n}(\lambda))\mathbf{y}\|^\mathbf{2}} (\mathbf{I_n} - \mathbf{A_n}(\lambda))\mathbf{y}, \\
\tilde{r}_n(\lambda) &= n^{-1}\|\tilde{\mathbf{f}}_n(\lambda) - \mathbf{f}\|^\mathbf{2}.
\end{aligned}$$

Theorem 2.3.6. *Under condition (A.2) and (A.3), $\hat{\mathbf{f}}_n(\hat{\lambda}_G)$ is consistent, i.e., $r_n(\hat{\lambda}_G) \to 0$, where $\hat{\lambda}_G$ is chosen by GCV.*

Theorem 2.3.7. *Under condition (A.1)-(A.3), $\hat{\mathbf{f}}_n(\hat{\lambda}_G)$ is asymptotically optimal, where $\hat{\lambda}_G$ is the GCV choice, i.e., we have $r_n(\hat{\lambda}_G)/r_n(\lambda_n^*) \xrightarrow{p} 1$, where λ_n^* is the best possible choice*

that is only known by oracles.

Theorem 2.3.6 establishes the asymptotic consistency of GCV's choice of the tuning parameter. Theorem 2.3.7 verifies the asymptotic optimality of the GCV. The proofs can be found in Appendix B.18 and B.19, respectively.

2.4 Discussion

The high-order Laplacian regularization studied in this work explores one direction of the Laplacian related regularizers. Other works [45, 46, 47, 49, 50, 51], which study the Laplacian related regularization, proceed with different perspectives compared to the high-order Laplacian regularization. Below, we summarize the related literature and state their differences from our work.

Johnson and Zhang (2007) [45], and Ando and Zhang (2007) [46] study the effect of Laplacian normalization in graph-based semi-supervised learning in the multi-class transductive learning, where the labels of the nodes in the graph are considered to be deterministic. Huang, Ma, Li, and Zhang (2011) [47] study a linear model, which belongs to the parametric estimation framework. They propose the sparse Laplacian shrinkage estimator (SLS), where they explicitly incorporate the correlation structure within the predictors (by using the information on the covariance matrix in the predictors) into the variable selection procedure and propose the Laplacian quadratic high-dimensional sparse estimation problems. The focus of SLS is the variable selection in high dimensional data, and they prove that the oracle property, meaning that it is sign consistent and equal to the oracle Laplacian shrinkage estimator, holds with high probability with the least squares loss function. Kirichenko and Zanten (2017) [50] investigate Bayesian regularization approaches and consider two types of priors on functions on graphs in a Bayesian framework. Kirichenko and Zanten (2018) [51] study the minimax lower bounds for function estimation problems on large graph, which relies on an assumption on their "asymptotic geometry". Both [50, 51] provide theoretical insights on Laplacian-related regularization

problems from the Bayesian and frequentists' point of view on large graphs, where they use the Laplacian matrix of large graph derived from the adjacency matrix. While in the current work, we investigate the properties of the high-order Laplacian regularization in a Sobolev space. We use the Laplacian matrix derived from the heat kernel. We analyze the eigenvalues of the discrete Laplacian matrix without relying on the asymptotic geometry assumption used in the aforementioned works.

2.5 Conclusion

In this work, we establish the theoretical foundation of the high-order Laplacian regularization in functional estimation. More specifically, we prove that under some realistic regularization conditions, the GLS estimator achieves the optimal convergence rate. To achieve our result, we study the asymptotic behavior of the eigenvalues of the gram matrix via making the connection between the semi-norm in Sobolev space and the spectrum analysis of the elliptic operator. In addition, we show that the generalized cross validation (GCV) still provides the best tuning parameter in terms of consistency and some form of asymptotic optimality. Through our description of the problem, we justify that the GLS estimator can be viewed as the generalization of both the thin-plate splines and the soap film smoothing (to the higher order Laplacian regularization) with nice theoretical guarantees.

CHAPTER 3

OPTIMAL SHAPE CONTROL VIA L_∞ LOSS FOR COMPOSITE FUSELAGE ASSEMBLY

3.1 Introduction

Recently, composite materials are widely used in large space structures due to its superior properties such as high strength-to-weight ratio. For example, an airplane of Boeing 787 comprises more than 50% composite parts by weight, and the fuselage is one key composite part of an aircraft [63]. In practice, there are natural dimensional variations in the fabrication of a fuselage due to different manufacturing batches or different suppliers [64]. When two fuselages assemble together, there is a gap, as shown in Figure 3.1. Thus, dimension variations influence the speed and the quality of composite fuselage assembly. As a result, the interface between two fuselages requires shape adjustments before the composite fuselage assembly process.

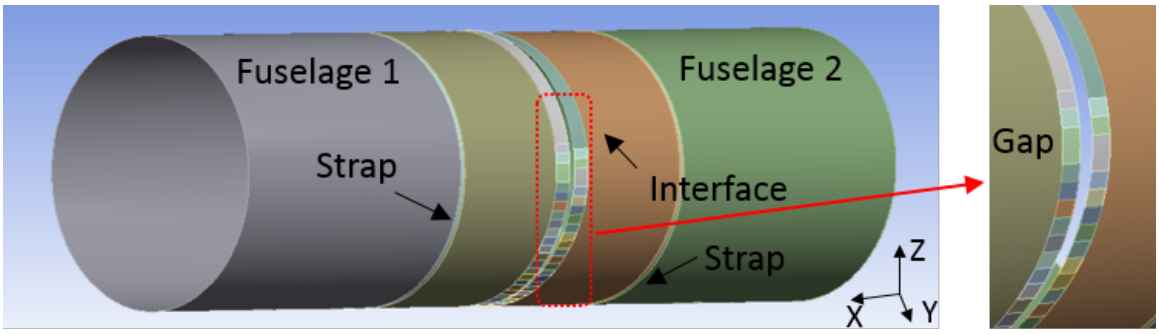

Figure 3.1: An illustrations of composite fuselage assembly.

In practice, actuators are used for shape adjustments of the interface between two composite fuselages, as shown in Figure 3.2. More details about using actuators for shape adjustments of composite fuselages can be found in [65], [66]. In this work, we focus on the shape adjustments of interface that are close to the edge plane of the fuselage. In the

state of the art, the target shape after control is the design shape, as shown in Figure 3.3. The dashed line is the design shape and the solid line is the shape of an incoming fuselage. The arrows show the actuators used for shape control of the interface. The current shape control strategy regards the target shape after adjustments as the design shape in terms of the ℓ_2 loss, which has two key limitations. (1) It is non-optimal. For a given pair of incoming fuselages with specific shapes, adjustment to the design shape is not optimal. In other words, there is no guarantee that the optimal shape control can be realized by shape adjustments to the design shape for two different incoming fuselages. Here, optimality means to achieve the minimum maximum gap on the interface between two fuselages after shape control. (2) For fuselage assembly, the maximum gap between two fuselages is the key issue preventing two fuselages assembly. Thus, the ℓ_∞ loss should be considered.

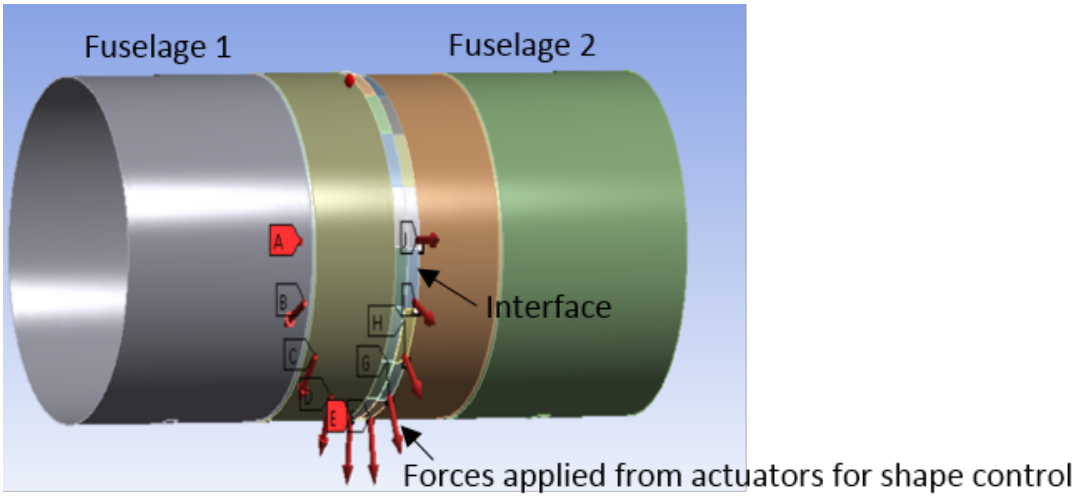

Figure 3.2: An illustration of shape adjustment by using actuators.

In the literature, multiple efforts are made to achieve shape control for structures. To focus on the research work related to this work, we mainly introduce the literature for composite fuselage shape control.

For the composite fuselage shape control, Wen et al. [65] first developed a new shape control system based on Finite Element Analysis (FEA) to improve the dimensional quality and productivity. Also, they show the feasibility of shape control for composite fuselages

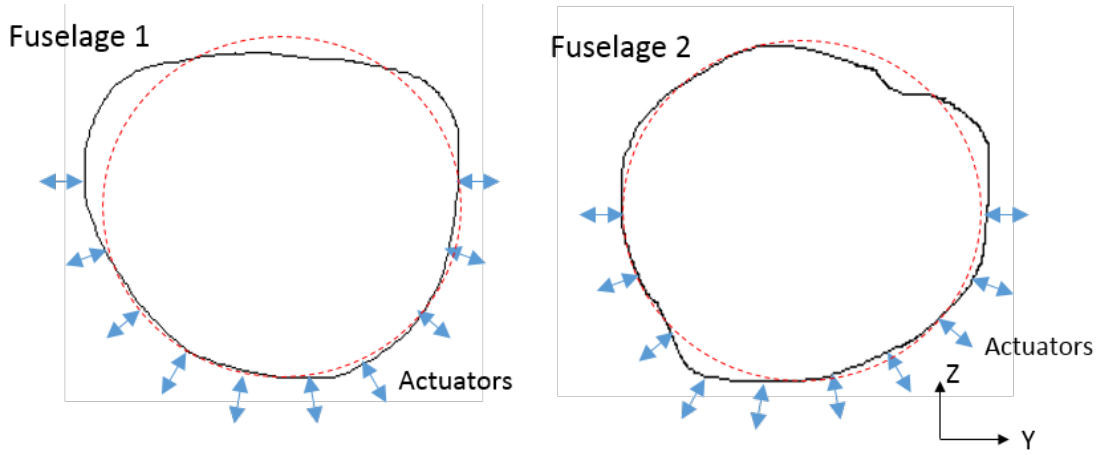

Figure 3.3: Schematics of interfaces between two fuselages. The dashed line is the design shape and the solid line is the shape of the interface of two fuselages. The arrows show the actuators used for shape control of the interface.

by using the proposed FEA platform. Based on this FEA platform, Yue et al. [66] proposed a surrogate model-based control strategy by considering the uncertainties for composite fuselage assembly. By minimizing ℓ_2 loss of dimensional errors compared to the design shape, they can calculate the optimal forces applied from actuators for single fuselage shape control. Based on the surrogate model, [67] proposed an optimal actuator placement strategy for the shape adjustment of composite fuselages, which reduces the forces applied from actuators and also dimensional deviations after shape control. However, all these works only consider shape adjustments of a single fuselage, and aim to adjust the incoming fuselages to the design shapes in terms of ℓ_2 loss, which meets the requirement of a single shape adjustment. However, when the two fuselages assemble together, the maximum gap, i.e., the ℓ_∞ norm of the gap, is one of the most important concern for fuselage assembly process. In terms of the maximum gap reduction for composite fuselage assembly, adjusting incoming fuselages into design shapes cannot guarantee the optimum given a specific pair of fuselages with different dimensional variations. Hence, an optimal shape control strategy is urgently needed for the composite fuselage assembly process.

To fill the research gap, this work proposes an optimal shape control strategy for composite fuselage assembly, which considers to minimize the maximum of dimensional gap

between a pair of incoming fuselages after control. Instead of adjusting each fuselage into the design shape, we consider the initial gap between the pair of fuselages and optimize the adjustment into an intermediate shape. As shown in [67], one other direct result is that optimal placement of actuators can also lead to the improvement of the shape control performance. In the current work, besides the focus on optimal shape control strategy for composite fuselage assembly, we also consider the optimal placement of actuators for two fuselage assembly. Given the dimensions of the pair of incoming fuselages, a sparse learning model is proposed to link actuator forces with the weighted maximum gap deviation of the pair of fuselages. In this way, the nonzero components of the force vector imply the optimal actuator locations. Often in practice, the last step is to refit the model with only the selected locations and get a more accurate estimation on the unknown forces. Hence, the optimal forces applied for the pair of fuselages can be obtained by minimizing the weighted maximum gap deviation.

In this work, we propose to use the ℓ_1 sparsity penalized ℓ_∞ loss linear regression to solve the composite fuselage assembly process. Specifically, we contribute to analyze the properties of the resulting estimation from our proposed penalized ℓ_∞ model under linear model with light tailed sub-Gaussian errors assumption, in the sparse estimation scenarios, and we provide the non-asymptotic upper-bound of the estimation error, measurement error, etc.; practically, we conduct case studies using our motivating example of composite fuselage assembly process to verify the effectiveness of our proposed procedure.

3.1.1 Notations

Throughout this work, we will need the following notations. We will denote vectors/matrix as Y, X, β, ϵ, etc. For a matrix $A \in \mathbb{R}^{m\times n}$, A_{ij} denotes the element of A at the (i,j)th location, and A_i denotes the $i$the column of A. Similarly, for a vector $\beta \in \mathbb{R}^m$, β_i denotes the ith element of β. For a subset $S \subset \{1, \cdots, m\}$, A_S denotes sub-matrix of A containing the columns with index in S, and β_S denotes sub-vector of β containing the elements with

index in S. We will use $\|\cdot\|_p$ to indicate the ℓ_p norm. Specifically, for a vector $x \in \mathbb{R}^n$, the ℓ_p norm of x is defined as: $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$. According to this definition, the ℓ_∞ norm of x is simply $\|x\|_\infty = \max_{i=1}^n |x_i|$. In case of the Euclidean norm, which is ℓ_2 norm, we will simply omit the subscript p, $\|\cdot\|$. We will denote the entrywise max norm of a matrix $A \in \mathbb{R}^{m\times n}$ as $\|A\|_\infty = \max_{1\le i\le m, 1\le j\le n} \|A_{ij}\|$. Throughout this work, we use the $sign(\cdot)$ function to indicate the sign of a vector or scalar. For $x \in \mathbb{R}$, $sign(x) = 1$ if $x > 0$ and $sign(x) = -1$ otherwise. For a vector $\beta \in \mathbb{R}^m$, $sign(\beta)$ is a vector with the ith element equal $sign(\beta_i)$. For a vector $\beta \in \mathbb{R}^m$, the support of β is defined as $supp(\beta) = \{i : \beta_i \neq 0\}$. And $|supp(\beta)|$ denote the cardinality of the set $supp(\beta)$.

3.1.2 Outline

In this work, we first provide a detailed description of the physical model of concern, including our justifications of using linear model with ℓ_∞ loss function in Section 3.2. Then we mathematically formulate our proposed model and show our main theoretical results in Section 3.3. Our main results use notations in statistics convention instead of the physics notations. A case study on fuselage assembly process using the FEA generated data is studied in Section 3.4, which verifies the effectiveness of our proposed model in engineering problems. Finally, we conclude our work in Section 3.5. All the proofs of our main theorems are postponed to the Appendix.

3.2 Fuselage assembly model

In this section, before going to the statistical analysis, we will provide necessary background of our physical model formulation. For the composite fuselage shape control, only elastic deformation is allowed during the shape adjustment. Thus, in this work, we assume linear mechanical behavior of fuselage deformation corresponds to the actuator forces.

Hence, the adjusted shape deviations can be formulated as

$$\delta_i = \psi_i + U_i F_i, i = 1, 2. \tag{3.2.1}$$

where $\delta_i \in \mathbb{R}_i^{2n}$ is the error in Y and Z directions after shape control and n denotes the number of measurement points for each fuselage; ψ_i and $U_i \in \mathbb{R}^{2n \times m_i}$ represent dimension deviations and displacement matrix of incoming fuselage i, respectively, and m_i denote the number of feasible positions (e.g., candidate positions where an actuator may be placed) for actuators in the i^{th} fuselage edge plane; F_i is the applied force during shape control of the fuselage i. The physical interpretation of the displacement matrix U_i is the deformation of all the measurement points correspond to the unit force on the structure. Without loss of generality, we assume the locations of measurement points are the same. Then, the dimensional gap between two fuselages after shape adjustment can be written as

$$\Delta = \delta_2 - \delta_1 = \psi_2 + U_2 F_2 - (\psi_1 + U_1 F_1). \tag{3.2.2}$$

Notably, the registration between two fuselages is needed if the number of measurement points for the pair of fuselages is not the same, which can be easily solved through [68].

For the composite fuselage assembly process, the main concern before assembly is the maximum gap point along the interface between the pair of fuselages. Thus, in this work, our objective is to minimize the weighted maximum gap between the pair of fuselages in both Y and Z directions, i.e., Δ_{max}, which is defined as

$$\Delta_{max} = \|B\Delta\|_\infty. \tag{3.2.3}$$

$B \in \mathbb{R}^{n \times n}$ is a diagonal weighted matrix, which represents the importance of the gaps in different measurement points. Such weight matrix is determined from the engineering domain knowledge. For example, we emphasize the dimensional gap on the upper fuselage

is more important, then we can add more weight in B matrix corresponding to the upper fuselage.

Usually in practice, we only have limited number of actuators available, from many potential positions for actuators, we will need to find the most effective ones, which means a sparse solution to our problem is desired. In order to encourage the sparsity in the resulting solution, we add a sparsity-induced penalty, namely, the ℓ_1 penalty, on the unknown forces at different locations:

$$\min_{F_1,F_2} J := \|B\Delta\|_\infty + \lambda\|F_1\|_1 + \lambda\|F_2\|_1, \tag{3.2.4}$$

where λ is the tuning parameter, which controls the sparsity of the resulting estimation.

From the statistical perspective, the above formulations can be viewed as the regression problem with ℓ_∞ loss and ℓ_1 regularization, and then the refitting procedure using only the chosen locations will help to reduce the bias on the estimation of the forces, F_1 and F_2, and result in a better estimation in practice.

In the statistics literature, the parameter estimation based on ℓ_∞ has been used in literatures since 1980, such as applications in the physical and environmental sciences [69, 70, 71], signal processing and systems engineering [72, 73, 74], etc. However, the theoretical guarantee is very limited, especially in the sparse estimation problems. There are two main directions studying regression analysis in the statistical literature.

The first one focuses on the asymptotic distribution in linear regression in a well-posed setting, where we have many more observations than the number of feature variables and there is no assumption on the sparsity of the unknown signals. [75] studied the problem of ℓ_∞ (or Chebyshev) estimator minimizes the maximum absolute residual under the situations where the noise distribution is known to have bounded support and unbounded support with light tails. They derived the asymptotic distribution of the estimator in the low dimension scenario.

The second one focuses on problems in the ill-posed settings, or settings where the unknown signal is sparse, where there are far more feature variables than the number of observations or there are a lot of 0 components in the signal. The Dantzig selector proposed in [12] solves the following problem:

$$\begin{aligned} &\min_{\beta \in \mathbb{R}^p} \|\beta\|_1 \\ &s.t. \|X^T(Y - X\beta)\|_\infty \leq \lambda\sigma, \end{aligned} \tag{3.2.5}$$

where the constrained optimization problem seeks to minimize the ℓ_1 sparsity objective function within the feasible region, where $\|X^T(Y - X\beta)\|_\infty \leq \lambda$. It can be easily seen that the term $X^T(Y - X\beta)$ is simply the first order derivative of the least square loss. Thus, the constraint $\|X^T(Y - X\beta)\|_\infty \leq \lambda$ will ensure that the scale of the first order derivative of the least square loss is very small. This basically, solves a similar problem as in lasso (minimizing the least square loss) [76].

However, theories to the previous models do not apply to our problem, where we have the incentive to minimize $\|Y - X\beta\|_\infty$ loss. In Section 3.3, we rewrite the above sparse estimation problem using ℓ_∞ loss in the statistical language and provide theoretical results on the non-asymptotic upper-bound of the estimation error, measurement error, etc., under the linear model setting with light tailed sub-Gaussian errors assumption.

3.3 Statistical model

In this section, we re-write the above physics model in statistics language to make the problem clearer with reader with no engineering background. Specifically, we give the statistical formulation in Subsection 3.3.1. Then we present our main theoretical results in Subsection 3.3.2.

3.3.1 Model in statistics language

We will consider the linear model throughout this work. Specifically, we assume that we observe data from the following model:

$$y = x^T\beta + \epsilon, \tag{3.3.1}$$

where y is the response variable, which in our case, is simply the shape deviation; x is the measurement vector; β is the unknown vector of forces we want to estimate and use on the actuators; ϵ is a random error. For all observations, it is assumed that ϵ_i, for $i = 1, \cdots, n$, are independent sub-Gaussian random variables with common variance parameter σ^2 and expectation 0. Let $Y \in \mathbb{R}^n$ denote the vector of response variable, $X \in \mathbb{R}^{n\times p}$ denote the observed measurement matrix. We will also use ϵ as a vector when the context is clear. We can write our model in the following matrix form:

$$Y = X^T\beta + \epsilon. \tag{3.3.2}$$

As in our motivated example, we want to estimate the forces which will minimize the ℓ_∞ norm of the adjusted shape deviation vector. We thus propose the following formulation for the estimation problem:

$$\min \|Y - X\beta\|_\infty, \tag{3.3.3}$$

where $Y - X\beta$ is the estimated adjusted shape deviation vector. However, any real world problem subjects to some unknown amount of noises, it usually is impossible to recover the exact signal as in [77]. Besides, in our motivated problem, we believe the underlying forces are sparse. Thus in this work, we mainly focus on the high-dimensional scenario where $p \gg n$, and the signal is sparse to guarantee the feasibility of the system. Without the sparsity assumption, the whole system is underdetermined. In order to drive the sparsity of the resulting solution, we will consider the regularized version of Problem (3.3.3) with

ℓ_1 penalty on the unknown parameter as follows,

$$\begin{aligned} &\min_{\beta \in \mathbb{R}^p} \|\beta\|_1 \\ &s.t. \frac{1}{\sqrt{n}} \|Y - X\beta\|_\infty \leq \lambda, \end{aligned} \tag{3.3.4}$$

where $\lambda > 0$ is a tuning parameter to control the sparsity of the resulting estimation, which we need to choose carefully, in order to recover the underlying ground truth. It can be easily checked that Problem (3.3.4) is a convex system and can be recast an a linear program (LP):

$$\begin{aligned} &\min_{\gamma \in \mathbb{R}^{+p}} \sum_{i=1}^{p} \gamma_i \\ &s.t. -\gamma_i \leq \beta_i \leq \gamma_i, \text{ for } i = 1, \cdots, p \\ &\quad -\lambda \leq \frac{1}{\sqrt{n}} (Y - X\beta)_j \leq \lambda, \text{ for } j - 1, \cdots, n. \end{aligned} \tag{3.3.5}$$

Problem (3.3.4) can also be equivalently written as the following Lagrangian form with some λ^*:

$$\min \frac{1}{\sqrt{n}} \|Y - X\beta\|_\infty + \lambda^* \|\beta\|_1. \tag{3.3.6}$$

For the interest of our proof, we use the constrained version as shown in Problem (3.3.4) for our theoretical analysis. We use the Lagrangian Formulation (3.3.6) for our numerical studies.

3.3.2 Main results

In this section, we present our main theoretical results on the estimation error of the resulting estimator proposed in Formulation (3.3.4). The proofs are postponed to the Appendix.

Restricted eigenvalue assumption

In this subsection, we will summarize the assumptions we need for deriving the main results. These are standard assumptions used in the statistical literature on recovering the

sparse signals.

Definition 3.3.1. *The restricted strong convexity (RSC) condition on model matrix X with respect to $\mathcal{C}$ is the following, there exists some constant $\gamma > 0$ such that:*

$$\frac{\frac{1}{n}\nu X^T X \nu}{\|\nu\|_2^2} \geq \gamma \text{ for all nonzero } \nu \in \mathcal{C}$$

here γ is called the restricted eigenvalue bound with regard to $\mathcal{C}$.

Assumption 3.3.1. *The restricted eigenvalues (RE) condition holds on the following set:*

$$\mathcal{C} = \left\{\nu \in \mathbb{R}^p \,|\, \|\nu_{\mathcal{S}^C}\|_1 \leq \|\nu_{\mathcal{S}}\|_1\right\}.$$

We have $\mathcal{C} \subset \mathbb{R}^p$ strictly since it is of the form of a cone.

The RSC (Assumption 3.3.1) is a standard assumption in the literature for proving the consistency results of regularized high-dimensional sparse estimation problems.

Estimation error upper-bound

Before we give the estimation error upper-bound, first we prove a lemma, which states that the unknown ground truth β^*, is feasible to Formulation (3.3.4) with high probability.

Lemma 3.3.1. *Let $\lambda = \sigma\sqrt{\alpha\frac{\log p}{n}}$ for some $\alpha > 2$, then with probability exceeding $1 - \frac{2n}{p} \cdot \left(\frac{1}{p}\right)^{\frac{\alpha-2}{2}}$, the unknown ground truth β^* is a feasible solution to Formulation (3.3.4).*

This lemma basically tells that, if the noise in the data generation is sub-Gaussian, when $\frac{2n}{p} \cdot \left(\frac{1}{p}\right)^{\frac{\alpha-2}{2}} \to 0$, the ground truth β^* lies in the feasible region of Formulation (3.3.4). Further, if the noise has a bounded distribution within the interval $[-\lambda, \lambda]$, β^* lies in the feasible region of Formulation (3.3.4) with probability 1.

Theorem 3.3.1. *Let $\hat{\beta}$ denote the optimal solution to Formulation (3.3.4). Suppose β^* is any unknown sparse ground truth with $|supp(\beta)| \leq S$, and the restricted eigenvalue*

assumption holds for the observed feature matrix X with γ in $\mathcal{C}$. Let $\lambda = O(\sigma\sqrt{\frac{\log p}{n}})$. We have the following non-asymptotic upper-bound on the estimation error:

$$\|\hat{\beta} - \beta^*\|_2 \leq O(\sigma\sqrt{S\log p}). \tag{3.3.7}$$

According to the above result, the performance of sparse parameter estimation from Formulation (3.3.4) is guaranteed, not only we will be able to recover the sparse unknown parameter in the high-dimensional scenario, we can bound the mean squared error (MES) of the resulting estimation in an order of a logarithmic factor to the true number of unknowns times the noise level, $O(\sigma\sqrt{S\log p})$. This upper-bound in fact is consistent with the MSE in [12].

However, we may obtain a better bound if in addition, the maximum absolute column sum of the matrix norm of the observed feature matrix X: $\|X\|_1 := \max_{i=1}^{p} \|X_i\|_1$ is also in the order of $\sqrt{n}$. This condition usually hold with sparse matrix, where there are a lot of 0 entries, or even matrix with columns, where the entry values decay exponentially. Specifically, we state our conclusion in the following theorem:

Theorem 3.3.2. *Let $\hat{\beta}$ denote the optimal solution to Formulation (3.3.4). Suppose β^* is any unknown sparse ground truth with $|supp(\beta)| \leq S$, and the restricted eigenvalue assumption holds for the observed feature matrix X with γ in $\mathcal{C}$. Let $\lambda = O(\sigma\sqrt{\frac{\log p}{n}})$. If we further have that $\|X\|_1 = O(\sqrt{n})$, we have the following tighter non-asymptotic upper-bound on the estimation error:*

$$\|\hat{\beta} - \beta^*\|_2 \leq O(\sigma\sqrt{\frac{S\log p}{n}}). \tag{3.3.8}$$

The proofs to Theorem 3.3.1 and 3.3.2 are in Appendix C.1.2 and C.1.3, respectively. The above Theorem 3.3.1 and 3.3.2 essentially states that with high probability, the optimal solution from our proposed formulation in (3.3.4) is very close to the unknown ground

truth, with their difference bounded by $O(\sigma\sqrt{\frac{S\log p}{n}})$.

According to [77, 12], the assumption on restricted eigenvalue condition will hold with high probability for random design matrices X, such as a random matrix with i.i.d. Gaussican entries, or Rademacher entries.

Prediction Error Upper-Bound

Theorem 3.3.3. *Under the assumptions of Theorem 3.3.2, we have the following upper-bound for the prediction error:*

$$\|X(\hat{\beta} - \beta^*)\|_2 \leq O(\sigma\sqrt{\frac{S\log p}{n}}). \tag{3.3.9}$$

The proof is simple, which is due to the by-product in the proof of Theorem 3.3.2 and is shown in Appendix C.1.4. This theorem tells that even if we are aiming to find a sparse solution to minimizing the ℓ_∞ loss function, the resulting solution will still produce a good estimation in the sense of mean squared error.

3.4 Case study

In this section, we study the fuselage assembly process using the linear model proposed in Section 3.2 with minor adjustment according to engineering requirements. We present our detailed experiment setting and report our results in Subsections 3.4.1 and 3.4.2, respectively.

3.4.1 Numerical setting

In this case study, we use an FEA model [65] to generate data and validate the proposed methodology due to the precious and limited data in the composite fuselage assembly process. The FEA model used in this case study has been validated with the experimental data, and more details about FEA model can be found in [65, 66]. For shape adjustments

of single fuselage, the method from [67] achieves the best control performance in terms of reduction of shape deviations after adjustments and the forces applied during shape adjustments.

Specifically, let m_1 and m_2 denote the numbers of all feasible locations for actuators along the pair of fuselages, while only a total number of M actuators are available for the shape control process in practice. Due to engineering specifications, and in order to meet the safety requirements during shape adjustment, we will solve a constrained problem (3.2.4), which is reformulated as follows:

$$\begin{aligned} &\min_{F_1,F_2} J := \|B\Delta\|_\infty + \lambda\|F_1\|_1 + \lambda\|F_2\|_1 \\ &s.t. F_{L_1} \leq F_1 \leq F_{Q_1}, F_{L_2} \leq F_2 \leq F_{Q_2}, \end{aligned} \tag{3.4.1}$$

Here, $F_{L1} \leq F_1 \leq F_{Q1}, F_{L2} \leq F_2 \leq F_{Q2}$ are the component-wise inequalities. F_{L1}, F_{L2} and F_{Q1}, F_{Q2} are the lower bounds and upper bounds of actuator forces for the first and second fuselages, respectively. Plugging in Equation (**??**), we have

$$\begin{aligned} &\min_{F_1,F_2} J := \|B(\psi_2 - \psi_1 + U_2F_2 - U_1F_1)\|_\infty + \lambda\|F_1\|_1 + \lambda\|F_2\|_1 \\ &s.t. F_{L_1} \leq F_1 \leq F_{Q_1}, F_{L_2} \leq F_2 \leq F_{Q_2}, \end{aligned} \tag{3.4.2}$$

Similarly as [67], $U_i, i = 1, 2,$ can be obtained from the surrogate model. From the solution of Problem (3.4.2), the optimal actuator placement is obtained from the support of F_1 and F_2 for shape adjustments of the pair of fuselages. In order to reduce the bias on the estimation of the forces, F_1 and F_2, we refit the model using the optimal positions and the optimal forces can be obtained by solving the following problem:

$$\begin{aligned} &\min_{F_1,F_2} J := \|B(\psi_2 - \psi_1 + (U_2)_{S_2}(F_2)_{S_2} - (U_1)_{S_1}(F_1)_{S_1})\|_\infty \\ &s.t. (F_{L_1})_{S_1} \leq (F_1)_{S_1} \leq (F_{Q_1})_{S_1}, (F_{L_2})_{S_2} \leq (F_2)_{S_2} \leq (F_{Q_2})_{S_2}, \end{aligned} \tag{3.4.3}$$

where $S_1 = supp(F_1)$, $S_2 = supp(F_2)$.

In order to make a fair comparison between our methods and the best result of composite fuselage shape control in current literature, we use the same FEA data from [67]. There are 20 incoming fuselages with different dimensional variations. For each pair of the fuselages, $m_1 = m_2 = 18$ feasible actuator locations distribute from -12 degrees to 192 degrees uniformly at the lower part of the fuselage from engineering practice. More details about the data generation can be refered to [67]. $M = M_1 + M_2 = 18$ actuators are used for shape adjustments of two fuselages in this case study. The number of measurement points along the interface of the pair of fuselages are $n_1 = n_2 = n = 182$, weight matrix $B = diag(1/n)$, and $U_1 = U_2$ are same with [67] for comparison purpose. The force bound is $F_{L_1} = F_{L_2} = -1000$ lbs and $F_{Q_1} = F_{Q_2} = 1000$ lbs. When we set F_{L_1} and F_{L_2} small enough, and F_{Q_1} and F_{Q_2} large enough, the solution to Problem (3.4.2) is the same as the unconstrained version, namely, Problem (3.2.4). This is true in the simulation study in the current work, which provides the theoretical guarantee of the optimal support choice according to Theorem 3.3.2.

In practice, some elements of displacement matrix $U_i, i = 1, 2$, are very small due to the structures of composite fuselages, such as the elements of U_i near the fixture. Specifically, we checked our simulated displacement matrix U_i and find that the element scales decays almost exponentially for each column, which satisfies our column norm assumption in Theorem 3.3.2. To achieve a better computational performance, such as reducing floating point errors [78], we multiply a large constant number L_N in the objective function $\|B(\psi_2 + U_2F_2 - (\psi_1 + U_1F_1))\|_\infty$ and $\|B(\psi_2 + U_2{}^cF_2{}^c - (\psi_1 + U_1{}^cF_1{}^c))\|_\infty$ in the optimization Problems (3.4.2) and (3.4.3), which do not have influences on the optimal solution. In this way, the computational problems induced by U_i matrix can be avoided in real implementations. In this case study, we set $L_N = 10^7$.

3.4.2 Results of the proposed method

In this case study, we randomly pick up two different fuselages from 20 incoming fuselages and have 50 replications. In this way, we have 50 pairs of fuselages for assembly. According to the engineering practice, we use root mean square gap (RMSG), maximum gap (MG), root mean square force (RMSF) and maximum force (MF) to evaluate the control results, which are defined as

$$RMSG := \frac{1}{n}\sqrt{\Delta' \cdot \Delta}, \tag{3.4.4}$$

$$MG := \sqrt{\|\Delta_Y * \Delta_Y + \Delta_Z * \Delta_Z\|_\infty}, \tag{3.4.5}$$

$$RMSF_i := \frac{1}{n}\sqrt{F_i' \cdot F_i}, i = 1, 2, \tag{3.4.6}$$

$$MF_i := \|F_i\|_\infty, i = 1, 2, \tag{3.4.7}$$

Here, '$*$' is the Hadamard product. $\Delta_Y \in \mathbb{R}^n$ and $\Delta_Z \in \mathbb{R}^n$ are the first and last n components of Δ, respectively. Thus, MG captures the maximum gap for two fuselage assembly. The control results of the proposed method on these 50 pairs of fuselages assembly are shown in Table **??**. Since in fuselage assebly process, we care more about the maximum gap after control and maximium force used for shape control, we also listed the maximum (max) of RMSG, MG and MF for these 50 pairs. The results show that the proposed method perfoms well in terms of the maximum gap by using relatively smaller forces, which is acceptable for fuselage assembly in practice. The following Figure 3.4 and Figure 3.5 show the box-plots of the fuselage gap after control and the maximum forces used for shape control in these 50 pairs.

Table 3.1: Control results of our method on 50 pairs of fuselages.

	RMG (inches)	MG (inches)	MF_1 (lbs)	MF_2 (lbs)
Mean	0.0014	0.0034	304.52	307.24
Max	0.0023	0.0060	553.83	768.03
Std	0.0004	0.0011	101.07	130.37

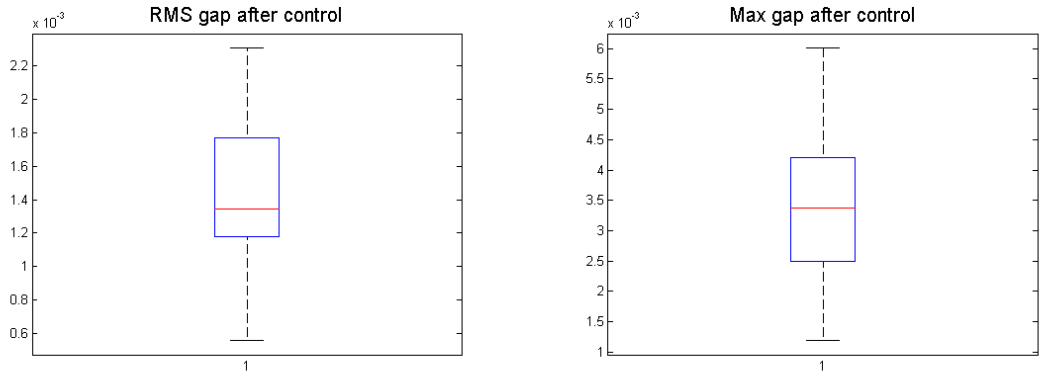

Figure 3.4: The boxplots of RMS gap and Max gap of 50 pairs of fuselages after control by the proposed method.

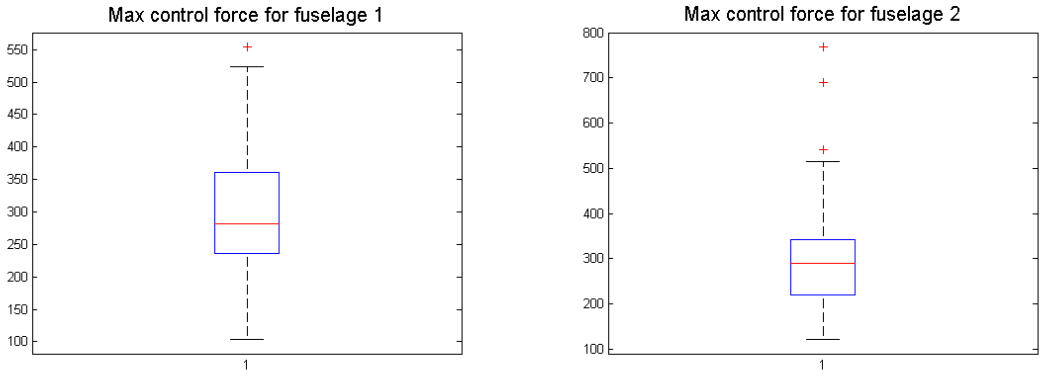

Figure 3.5: The boxplots of max control forces for fuselage 1 and fuselage 2 by the proposed method.

Comparisions

In the following, we compare the proposed method with the current practice [67], which is to control the incoming fuselage into design shape in terms of ℓ_2 loss, and then assemble. We calculated the improvements on the force and gap after shape adjustments, and use statistical t test to test the significance of the improvement on the gap and applied forces. The improvement results on mean and standard deviation (std) of the gap after adjustments are listed in Table **??**. Notably, the improvement means the control results via current practice minus the results of the proposed method. The null hypothesis H_0: the improvement comes from a distribution with mean zero, i.e., no significant improvement. The alternative hypothesis H_1: mean of improvement is greater than 0 (right-tailed test), which indicates the improvement of the proposed method. The $p-$value is also listed in Table **??**. As shown in Table **??**, compared to [67], the proposed method significantly improves the maximum gap (max gap) and RMS gap for two fuselage assembly. The box-plots of MG and RMSG improvement are shown in Figure 3.6. Since we achieve much smaller gap after control, the controlled forces are expected to increase. Thus, for the force comparison, the alternative hypothesis H_1: mean of improvement is less than 0 (left-tailed test). The results are listed in Table **??**. The box-plots of MF Improvement for Fuselage 1 and Fuselage 2 are shown in Figure 3.7. Although the results show that the proposed method uses about 60 lbs larger forces to control the fuselage shape, such amount increase is not a problem for fuselage shape control in practice. It is still in the safety region, the increase amount is very limited. From practice, it is regarded as comparable in terms of forces used for fuselage shape control compared to [67]. However, our method significantly improves the MG after shape control by applying the comparable forces for composite fuselage shape control. Notably, the mean of MG improvement is 0.0154 inches, which will significantly help improve composite fuselage assembly process.

Method	Max RMSG (inches)	Max MG (inches)	Max $RMSF_1$ (lbs)
Our method	0.0086	0.0134	174.8357
Method from [67]	0.0096	0.0349	156.1150
Method	Max MF_1 (lbs)	Max $RMSF_2$ (lbs)	Max MF_2 (lbs)
Our method	292.1276	175.1264	395.9643
Method from [67]	298.5735	200.7488	416.6345

Table 3.2: Gap reduction of our method compared with method from [67].

MG Improvement (inches)			RMSG Improvement (inches)		
Mean	Std	P-value	Mean	Std	P-value
0.0154	0.0074	6.17×10^{-20}	0.0048	0.0023	5.38×10^{-20}

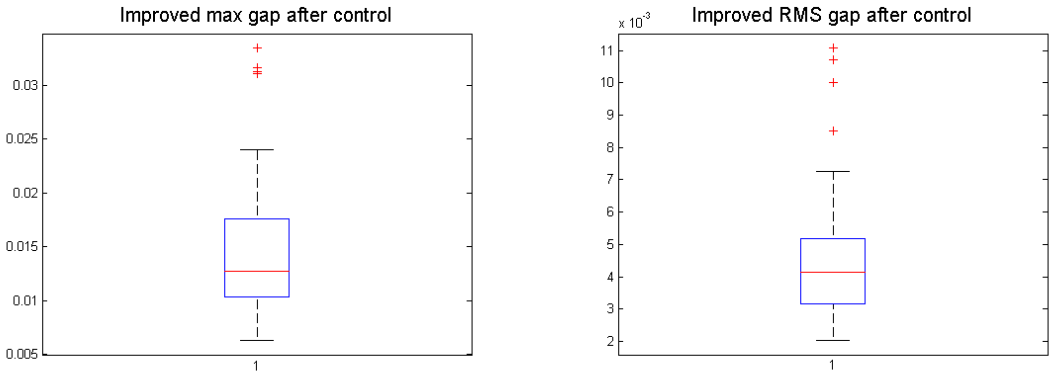

Figure 3.6: The boxplots of improved max and RMS gap after control compared with method from [67].

Table 3.3: Max force increase of our method compared with method from [67].

MF Increase for Fuselage I (lbs)			MF Increase for Fuselage II (lbs)		
Mean	Std	P-value	Mean	Std	P-value
53.50	103.57	3.16×10^{-4}	58.0696	156.96	0.0059

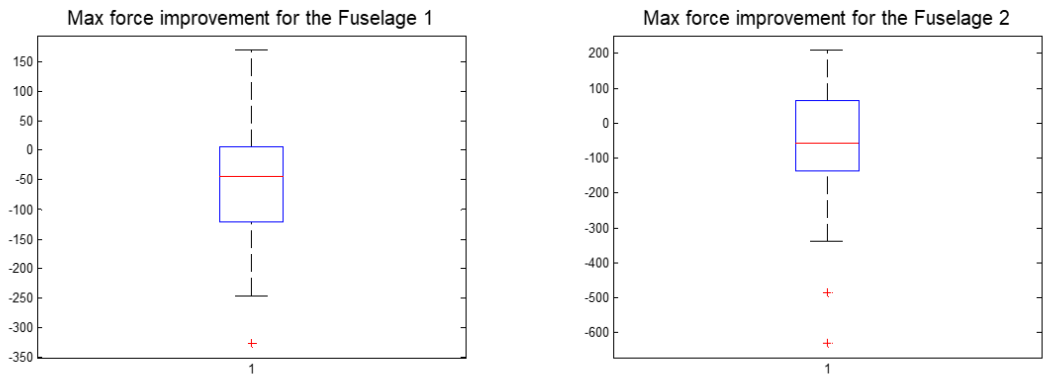

Figure 3.7: The boxplots of max force improvement for fuselage 1 and fuselage 2 of our method compared with method from [67].

3.4.3 Discussions

Notably, our goal of this work is to propose an optimal shape control strategy for composite fuselage assembly process. The main concern is the maximum dimensional gap during the composite fuselage assembly process in practice. Thus, the objective of this work is to minimize the maximum gap point for composite fuselage shape control instead of mean square error in [67]. The goal of [67] is to achieve the optimal actuator placement for shape control of single fuselage, and the target shape after shape control is the design shape. Hence, in the case study, our method performs much better than [67] in terms of the maximum gap after shape control. Also, thanks to considering the initial gap, the RMSG is also smaller than [67] by applying comparable forces.

3.5 Conclusions

This work proposes an optimal shape control strategy for composite fuselage assembly process. Due to natural dimensional variations of fuselages, there is a gap on the interface of two fuselages before assembly. The current practice adjusted the shape of each fuselage to the design shape in terms of ℓ_2 loss and then assemble, which is not optimal. Our contribution is to consider the initial gap of the pair of incoming fuselages and pro-

pose a sparse learning model, which aims to minimize the maximum gap (ℓ_∞ loss) after shape control. The proposed model is simply ℓ_1 sparsity penalized ℓ_∞ loss linear regression. Under linear model with light tailed sub-Gaussian errors assumption, we provide the non-asymptotic upper-bound of the estimation error, measurement error, etc.. Practically, we conduct case studies using our motivating example of composite fuselage assembly process. We show that our method uses small forces to achieve very small maximum gap after control. Compared to the current literature, a set of case studies show that our method achieves significant reduction of the maximum gap after shape control by applying comparable forces. Notably, although our method is demonstrated for optimal shape control in composite fuselage assembly process, the methodology can be extended for optimal shape control of assembly process in other structures.

CHAPTER 4

DISPARITIES IN ACCESS TO PREVENTIVE DENTAL CARE BETWEEN PUBLICLY AND PRIVATELY INSURED CHILDREN IN GEORGIA

4.1 Introduction

Children living in poverty are over twice as likely to have untreated tooth decay compared to children with family incomes $> 200\%$ of the federal poverty level (FPL; 25% versus 12%) [79]. Tooth decay if left untreated can lead to problems in eating, speaking, and learning [80]. There is strong evidence for the effectiveness of preventive dental services [81] and increasing low-income children's access to them is a national health goal [82]. A major barrier to poor children not receiving dental care is difficulty in finding a dentist who accepts Medicaid insurance [83]. Policies and programs aimed at increasing access to preventive dental care (e.g., increasing dental providers or providing services in schools) are typically implemented locally [84, 85, 86].

Building on previous research, we estimated three measures of local access – percentage of *met need* for preventive dental services, one-way *travel distance* to a dentist, and *dentist scarcity* – for children living in Georgia census tracts. We compared local access for two groups, children whose family income would qualify for public dental insurance (Medicaid or the Children's Health Insurance Program (CHIP)) and children living in families with high income (>400% FPL) which are assumed either to have private insurance or to be able to afford out-of pocket expenses. We also estimated these measures for rural and urban tracts. Finally, we examined the impact of increasing dental provider participation in Medicaid/CHIP on preventive dental care access in both groups.

4.2 Methods

4.2.1 Study population

We used data from the US Census and American Community Survey to compare access to preventive dental services for Georgia children, aged 0 to 18 years, living in households with family incomes 247% of the federal poverty level (income threshold for Medicaid/CHIP eligibility(9); hereon referred to as *publicly-insured children* to those with family incomes $>$ 400% of the federal poverty level; hereon referred to as *privately-insured children* (Web-Appendix Section 1.1 [87]).

4.2.2 Access Measures

We calculated three measures of access for each census tract:

- *Percentage of met need* – total met need divided by pediatric need for preventive services. Met need refers to need served within the state access standards [88]. High values indicate smaller proportions of children who need to travel longer distances than the state access standards to reach an available provider.

- *Travel distance* – average distance in miles a child must travel one-way to visit the dentist. The travel distance is provided only for those children with travel distances within the state access standards [88]. High values indicate large travel distances. The travel distance is computed using a street network using the GIS ArcMap software.

- *Provider scarcity* – patient caseload served by dentists divided by maximum patient caseload capacity. High values indicate high scarcity of providers.

We also designated census tracts as *served*, *underserved* and *unserved* if the proportion of children with unmet need within the state access standards and uninsured children in

households without ability to afford dental care was $\leq 10\%$, $10\% - 50\%$, and $> 50\%$, respectively.

We estimated these measures across all census tracts in Georgia and across rural (located in counties with population $< 35,000$) and urban tracts (population $\geq 35,000$) [89]. Further detail on the calculations for these measures is provided below.

4.2.3 Estimating Need

To estimate need for pediatric preventive dental services, we used a published methodology [90] to estimate the number of dental provider hours required to provide preventive dental services at frequency recommended by the American Academy of Pediatric Dentistry [91, 92]. Recommended services and frequency of delivery depend on a child's age and risk status for caries. Further detail is provided in Web-Appendix Section 1 [87].

4.2.4 Estimation of Preventive Dental Care Supply

We obtained a list of Georgia dentists and their practice addresses from the 2015 Board of Dentistry. We used their taxonomy code (2015 National Plan and Provider Enumeration System) to identify providers of preventive services to children (further details in Web-Appendix Section 2 [87]). The addresses of individual providers were geocoded using the Texas A&M Geocoding Services [93]. Street-network distances between provider addresses and census tract centroids were computed using ArcGIS Network Analyst [94]. Maximum capacity for preventive dental care for children per provider is estimated following existing estimation procedures [90]. Note that the proportion of provider capacity allocated to prevention was based on the distribution of services as defined in the Medical Expenditure Panel Survey (MEPS) in units of time. Details are provided in Web-Appendix Section 3 [87].

In order to estimate the number of providers accepting Medicaid/CHIP (public insurance) in each census tract, we used data from InsureKidsNow.gov (IKN). Using an ap-

proach similar to the American Dental Association (ADA) [95], we matched providers recorded as accepting public insurance in the IKN database with providers in the Board of Dentistry list using fuzzy logic, after removing repeats in the IKN data and accounting for both individual providers and dental care offices. For dental care offices, we assumed all dentists identified in an office appearing in the IKN data accepted public insurance.

The 2012 MAX Medicaid claims data obtained from CMS were used to estimate the distribution of capacity allocated by each provider for publicly insured children accounting for excess capacity due to no-shows and potential underutilization. The Institutional Review Board protocol number is H11287. The detailed procedures to derive the supply estimates are described in Web-Appendix Section 2 [87].

4.2.5 Optimization Model

To estimate access, we used an optimization model [96] to match dental supply and need under the following set of constraints:

- *Supply availability*: number of patients assigned to each provider does not exceed maximum caseload capacity for pediatric preventive care (i.e., provider scarcity ≤ 1);
- *Public insurance acceptance*: number of assigned publicly-insured patients does not exceed provider's public insurance caseload;
- *Patient's travel mobility*: patients travel distance does not exceed Georgia guidelines on access standards [97]. The maximum distance for patients with personal vehicles is 30 miles in urban areas and 45 miles in rural areas [97]. For patients without a private vehicle who must use an alternative means of transportation, we set a maximum distance threshold of 15 miles (45 minutes travel time) for rural tracts and 8 miles (30 minutes travel time) for urban tracts.

Based on the assumption that patients prefer closer providers, the objective of the optimization model was to minimize total distance traveled to reach dental providers by

publicly-insured and privately-insured children. We did not include uninsured children from families with incomes between 247% and 400% directly in the optimization model since they are assumed without financial access. Details are provided in Web-Appendix Section 1 [87].

The model determined the number of children in the two study populations for each census tract assigned to a provider location based on the previously mentioned constraints. Need within a census tract could be assigned to different providers; if insufficient provider capacity, a proportion of need could be unmet. Because many providers do not accept Medicaid/CHIP patients, our model assigned privately-insured and publicly-insured children separately. In order to account for uncertainty in provider caseload's estimates and in proportion of high-risk children (i.e., higher need for preventive dental care) we ran 50 microsimulation runs simultaneously sampling from these parameters. Details and justification are provided in Web-Appendix Section 3 [87].

4.2.6 Disparities

A disparity was defined as the absolute difference in access between publicly-insured children (low family income) and privately-insured children (insured and/or with high family income). Using a simultaneous inference approach [98], we identified census tracts with access statistically worse than various disparity thresholds at the significance level 0.05. For travel distance, we tested the disparity thresholds of 2, 6, 8, and 10 miles. For provider scarcity, we tested the disparity thresholds of 0, 0.1, 0.2, and 0.3. Both thresholds were chosen based on the authors' belief they were reasonable ranges to consider. We showed the location of the census tracts where access is below than these disparity thresholds on point maps (i.e., significance maps) where the points in the maps correspond to census tract centroids for which access is statistically significantly worse than the disparity thresholds. Details are provided in Figure 4.1. Each dot on the map corresponds to a census tract where the publicly-insured population has a statistically significantly greater distance or greater

scarcity of providers than the privately-insured population, at $\alpha = 0.05$ significance level in at least 75% of the runs. The grey-shaded regions on the maps correspond to counties where the publicly-insured population does not experience a significantly worse accessibility or availability in at least than 75% of the runs. The map on the bottom shows the urbanicity classifications of census tracts in the state according to State Office of Rural Health. We group census tracts into two categories: Rural (counties with population less than 35,000) and Urban (otherwise) areas.

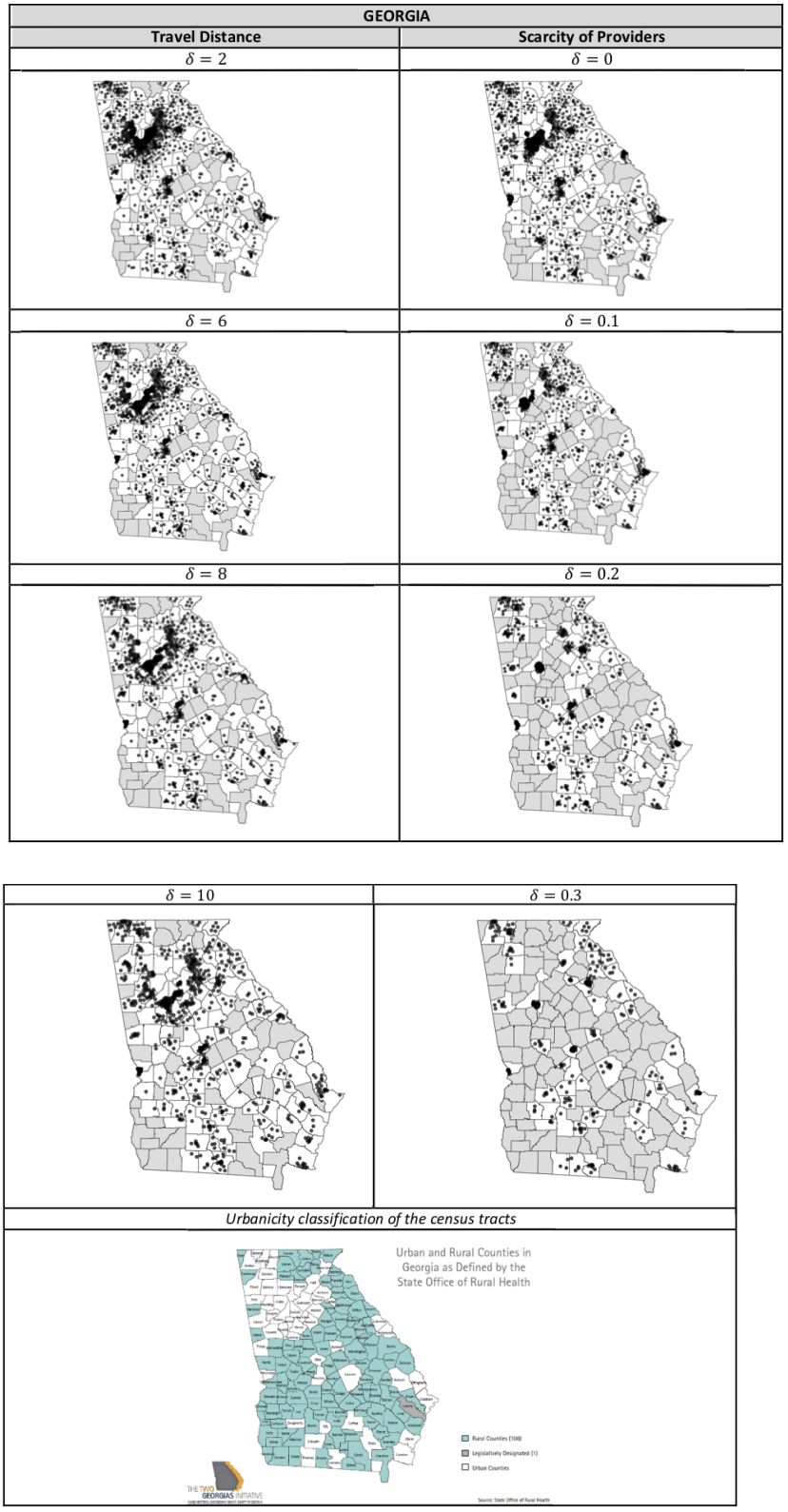

Figure 4.1: Significance maps.

4.2.7 Impact of Changing Dentist Participation in Medicaid on Access

We examined the impact of varying provider acceptance rates of public insurance for children from 20% to 80% on our three access measures. To do this, we first set the acceptance rate to a given value and then sampled the Medicaid caseload differently for providers in urban (caseload ranged from 35% to 50%) and rural (range: 55% to 65%) tracts. Similarly, we varied the capacity of providers accepting Medicaid patients from 20% to 75% and setting the maximum allowed travel distance to providers for families owning a vehicle from 30 to 60 miles.

4.3 Results

4.3.1 Study Population

Among the approximately 2.6 million Georgia children, the estimated number of publicly-insured children was 1.5 million and the number of privately-insured children 600,000. There were 1,969 census tracts (1527 urban) in 159 counties (50 urban). The number of publicly-insured children was 1,183,470 and 309,813 in urban and rural tracts, respectively. The number of privately-insured children was 536,043 and 68,194 in urban and rural counties, respectively. A map of the distribution of the percentage of children in these two groups over census tracts in Georgia is provided in Figure 4.2.

4.3.2 Overall Dental Supply and Access in Georgia

There were 4,123 dentists providing preventive dental care to children. Among these providers, 27.9% accepted public insurance (IKN database).

4.3.3 Access measures

The state-level average met need for publicly- and privately-insured children was 0.59 (10^{th} percentile=0.00, 90^{th} percentile=1.00) and 0.96 (0.99, 1.00), respectively (Table **??**). In

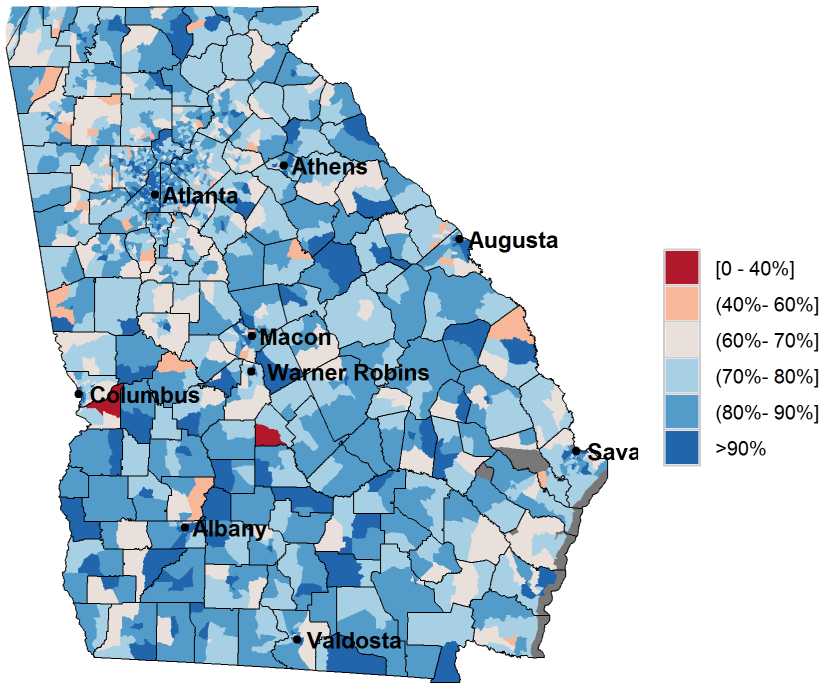

Figure 4.2: Financial access at the census tract level (percentage of children with financial access to preventive dental care at each census tract). Financial access is the percentage of children who either are eligible for public insurance or have ability to afford dental care through commercial insurance or ability to pay out-of-pocket. Location with low percentages of children with financial access are those that have a large percentage of children without ability to afford dental care.

rural areas, these values were 0.33 (0, 0.89) and 0.84 (0, 1), and in urban areas 0.67 (0, 1) and 0.99 (0.99, 1) for publicly- and privately-insured children, respectively. The average travel distance for publicly- and privately-insured children was 17.16 miles (1.11, 45.00) and 3.71 miles (0.02, 7.28), respectively. In rural areas, the average travel distance was 32.91 (10.31, 45) miles and 11.55 (0.61, 45) miles and in urban areas 12.62 (0.74, 30) miles and 1.46 (0.01, 3.56) miles for publicly- and privately-insured children, respectively. The average provider scarcity for publicly- and privately-insured children was 0.70 (0.39, 1.00) and 0.45 (0.05, 0.91), respectively. In rural areas, average provider scarcity was 0.88 (0.65, 1) and 0.50 (0.09, 1) and in urban areas 0.65 (0.38, 1) and 0.43 (0.04, 0.89) for publicly- and privately-insured children, respectively. Boxplots of these access measures

are provided in Figure 4.3.

Table 4.1: Results for average values (10^{th} percentile, 90^{th} percentile) of access measures across all 50 simulated settings, for publicly insured children and those privately insured children or high family income, for all census tracts and also differentiated for rural and urban tracts. Georgia 2015.

Level	Percentage of Met Need		
	Entire Population	Publicly Insured	Privately Insured
State	0.67 [0.14, 1]	0.59 [0, 1]	0.96 [0.99, 1]
Rural	0.42 [0, 0.92]	0.33 [0, 0.89]	0.84 [0, 1]
Urban	0.74 [0.26, 1]	0.67 [0, 1]	0.99 [0.99, 1]
Level	Travel Distance		
	Entire Population	Publicly Insured	Privately Insured
State	14.4 [0.52, 36.43]	17.16 [1.11, 45]	3.71 [0.02, 7.28]
Rural	29.26 [7.95, 45]	32.91 [10.31, 45]	11.55 [0.61, 45]
Urban	10.12 [0.34, 23.5]	12.62 [0.74, 30]	1.46 [0.01, 3.56]
Level	Scarcity of Providers		
	Entire Population	Publicly Insured	Privately Insured
State	0.67 [0.38, 0.95]	0.7 [0.39, 1]	0.45 [0.05, 0.91]
Rural	0.82 [0.57, 1]	0.88 [0.65, 1]	0.5 [0.09, 1]
Urban	0.63 [0.35, 0.91]	0.65 [0.38, 1]	0.43 [0.04, 0.89]

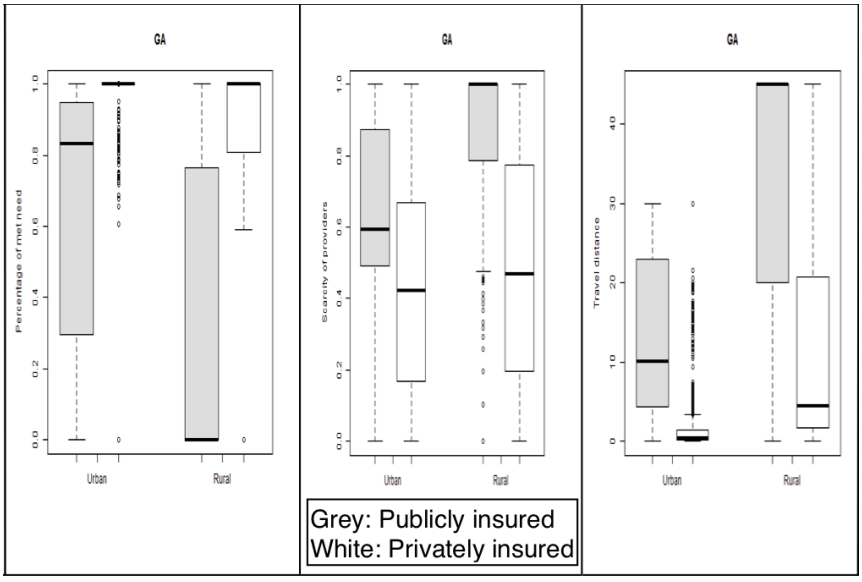

Figure 4.3: Boxplots of the distribution of the percentage of met need (left), travel distance (middle) and scarcity of providers measured at the census tract level for different geography and for the two population groups: publicly-insured and privately-insured population of children. Measures used here are the medians computed from 65 runs at the census tract level.

Assuming a capacity ranging between 35% and 65% for urban, and between 55% and 65% for rural communities, we found that 6% of the census tracts were served, 57% under-served, and 37% unserved (Table **??**).

Table 4.2: Average [minimum, maximum] percentage of census tracts in each service level category for all, urban and rural census tracts across the 65 simulations. Georgia 2015.

	% of tracts in service level category		
Level	Entire Population	Publicly Insured	Privately Insured
State	6 [2, 8]	57 [56, 60]	37 [36, 38]
Rural	8 [3, 10]	64 [62, 68]	29 [27, 31]
Urban	1 [0, 2]	35 [32, 38]	64 [61, 67]

4.3.4 Disparities

The difference in travel distance between publicly- and privately-insured children was greater than 2 miles for 72% of the census tracts, and greater than 10 miles for 38% of

the census tracts (Table **??**). The difference in provider scarcity was greater than 0 in 68% of census tracts and greater than 0.3 in 16% of census tracts.

Table 4.3: Results for number (%) of census tracts where absolute difference in access measure between the two child sub-populations for multiple met threshold criteria: Georgia 2015.

Level	Travel Distance			
	2 (miles)	6 (miles)	8 (miles)	10 (miles)
State	1399 (72%)	1104 (56%)	934 (48%)	749 (38%)
Rural	304 (22%)	262 (24%)	243 (26%)	219 (29%)
Urban	1095 (78%)	842 (76%)	691 (74%)	530 (71%)
Level	Scarcity of Providers			
	0	0.1	0.2	0.3
State	1321 (68%)	919 (47%)	612 (31%)	307 (16%)
Rural	312 (24%)	235 (26%)	161 (26%)	107 (35%)
Urban	1009 (76%)	684 (74%)	451 (74%)	200 (65%)

4.3.5 Impact of increased provider participation in Medicaid on access

Access increased among publicly-insured children as provider participation in Medicaid increased (Figure 4.4). For a provider participation of 20%, median met need was 30.5%, median travel distance was 5.56 miles, and provider scarcity was 0.86. Although not realistic in practice, in order to achieve 100% median met need, an 80% provided participation would be required. This would also result in a decrease median travel distance to 5.56 miles and provider scarcity of 0.52. For rural tracts, the median met need increased from 21.7% to 100%; the median travel distance decreased from 38.92 miles to 20.18 miles; and provider scarcity decreased from 0.94 to 0.65 for the same increase in provider participation increase. For urban tracts, the median met need increased from 46.7% to 100%; the median travel distance decreased from 19.15 miles to 3.80 miles; and provider scarcity decreased from 0.83 to 0.47.

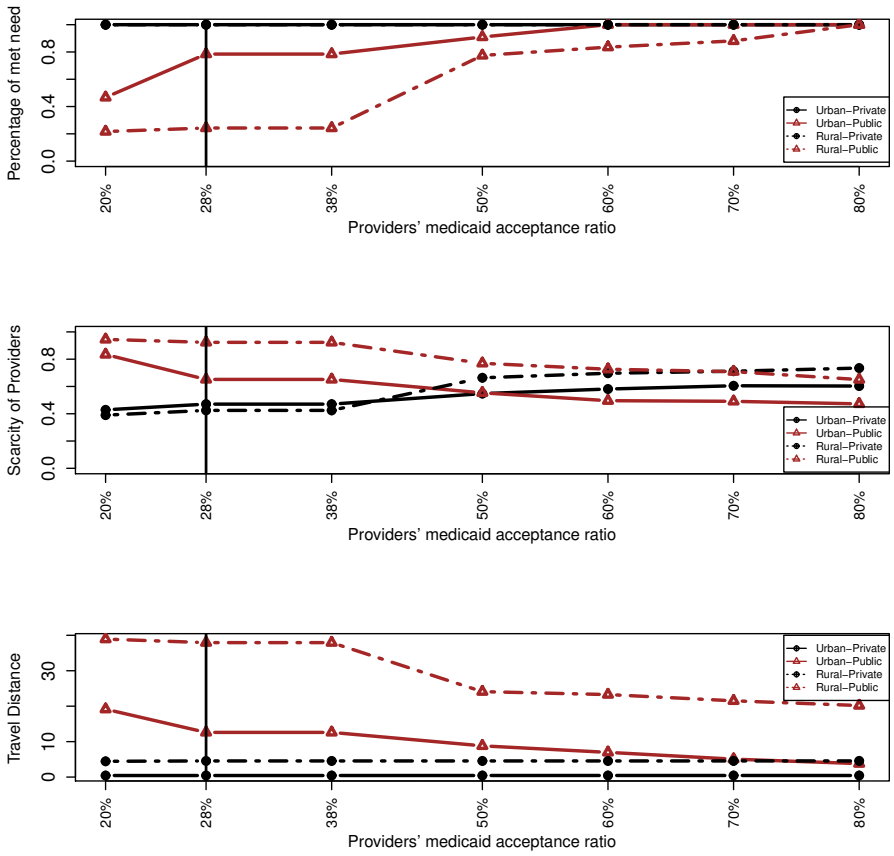

Figure 4.4: Median values of the percentage of met need, travel distance, and scarcity of dentists in rural and urban census tracts, by dentists' Medicaid/CHIP acceptance ratio. Scarcity was calculated as the patient caseload served by dentists divided by maximum patient caseload capacity; higher values indicate greater scarcity of dentists. The vertical dashed line at 28% represents the current rate of providers participating in public insurance programs. Abbreviation: CHIP, Children's Health Insurance Program.

For privately-insured children, the effect was negligible – median met need is 100%; median travel distance increased from 0.71 to 0.74 miles; and provider scarcity increased from 0.42 to 0.65 at the state level.

Holding other variables constant, increasing the Medicaid caseload of providers currently accepting Medicaid patients from 20% to 75% also increased access among publicly-insured children according to the optimization model results. Met need increased from 24.0% to 98.1% at the state level, from 17.4% to 79.6% for rural tracts and from 26.2%

to 100% for urban tracts. At the state level, travel distance decreased from 24.81 to 10.42 miles; and provider scarcity decreased from 0.85 to 0.70 (Figure 4.5). For privately-insured children, the effect was again negligible – median met need is 100%; median travel distance increased from 0.71 to 0.77 miles; and provider scarcity increased from 0.36 to 0.54. We show the results at rural and urban census levels in Figures 4.6 and 4.7.

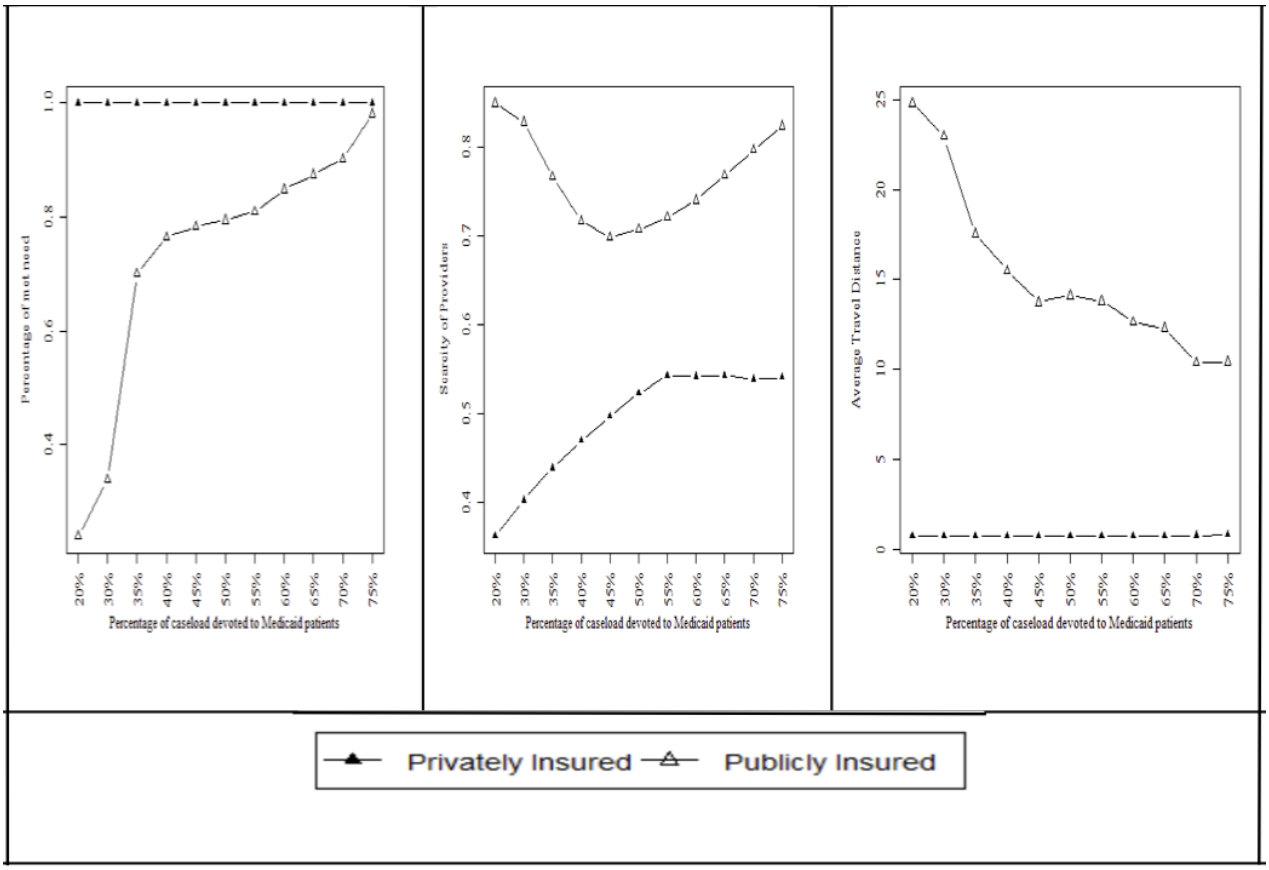

Figure 4.5: Sensitivity analysis of the percentage of met need, travel distance and scarcity of providers at the state level with respect to changes in percentage of providers' caseload devoted to publicly insured patients.

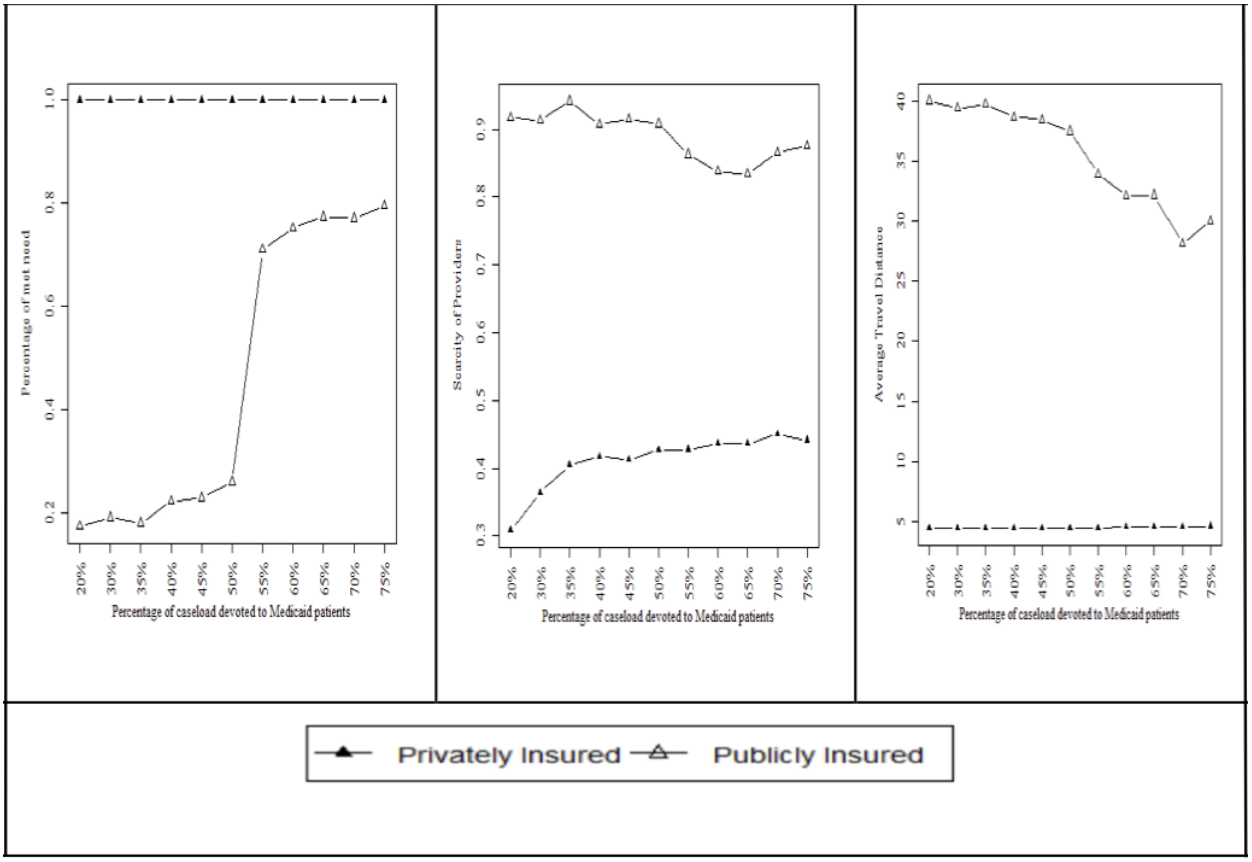

Figure 4.6: Sensitivity analysis of the percentage of met need, travel distance and scarcity of providers for rural areas with respect to changes in percentage of providers' caseload devoted to publicly insured patients.

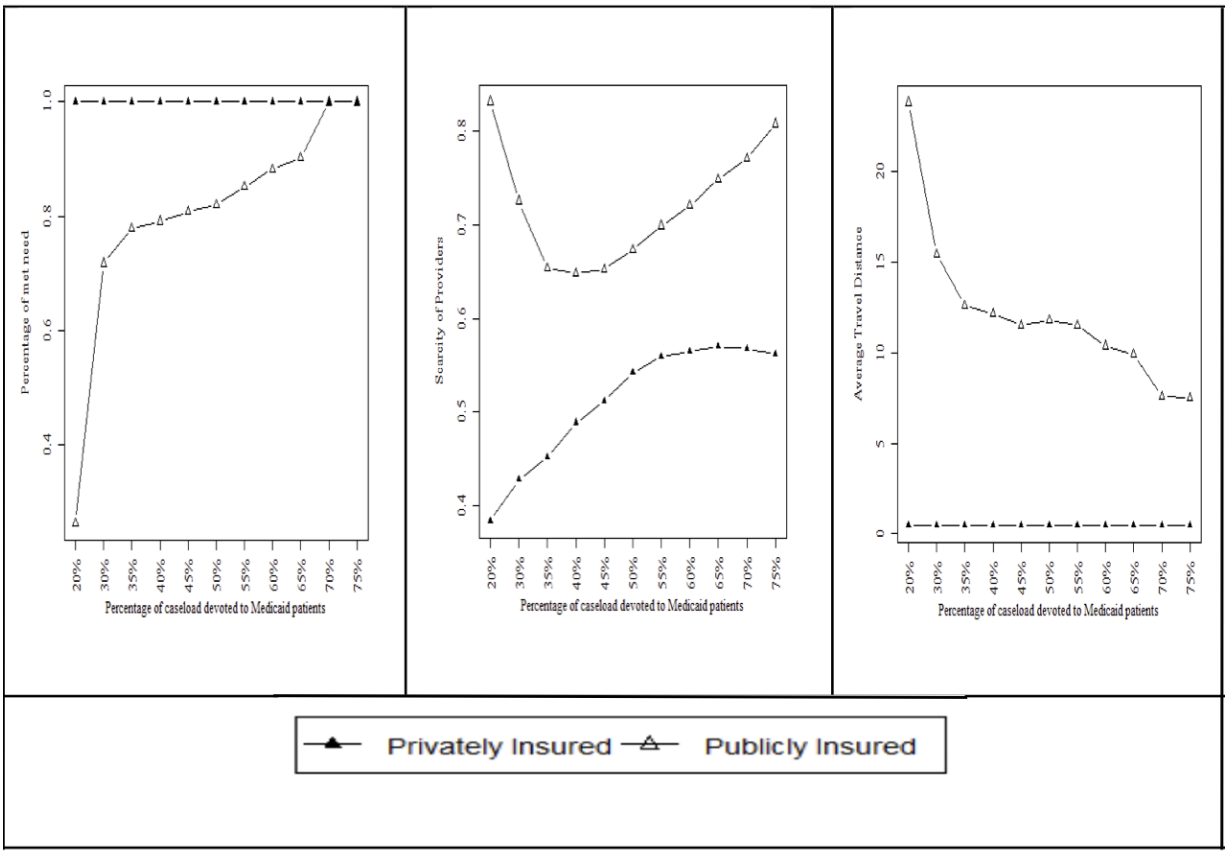

Figure 4.7: Sensitivity analysis of the percentage of met need, travel distance and scarcity of providers for urban areas with respect to changes in percentage of providers' caseload devoted to publicly insured patients.

Figures 4.8, 4.9, 4.10 shows how the access measures vary for maximum allowed travel distance for people with a personal vehicle from 30 miles to 60 miles at state level for urban and rural communities. At the state level, percentage of met need stayed at 76.4% for children eligible for public insurance, and stayed at 100% for children from higher income families; state-level median travel distance varied between 15.13-25.51 miles for children eligible for public insurance, and stayed at 0.71 miles for children from higher income families; state-level median provider scarcity stayed around 0.73 for children eligible for public insurance and between 0.46-0.47 for children from higher income families.

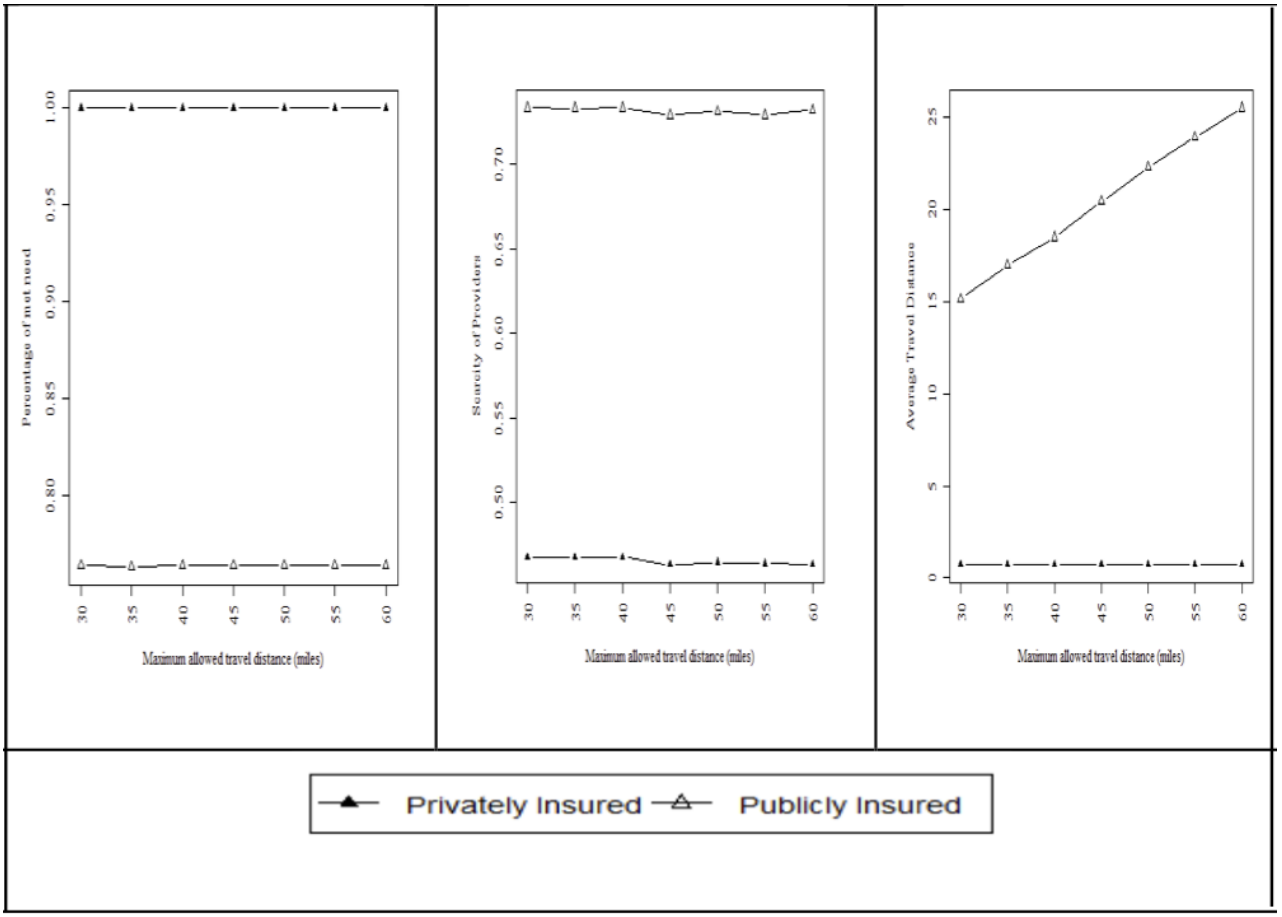

Figure 4.8: Sensitivity analysis of the percentage of met need, travel distance and scarcity of providers at the state level with respect to changes in maximum allowed travel distance parameter.

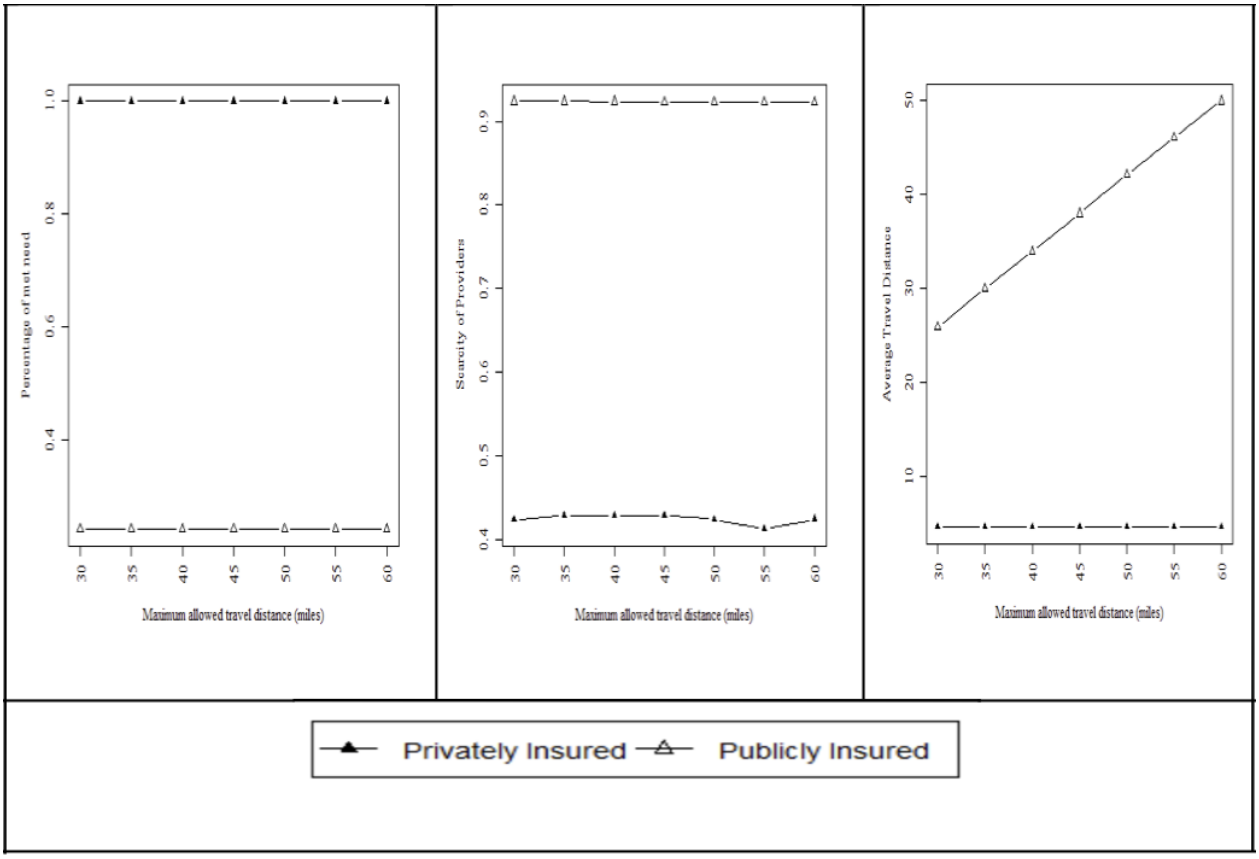

Figure 4.9: Sensitivity analysis of the percentage of met need, travel distance and scarcity of providers for rural areas with respect to changes in maximum allowed travel distance parameter.

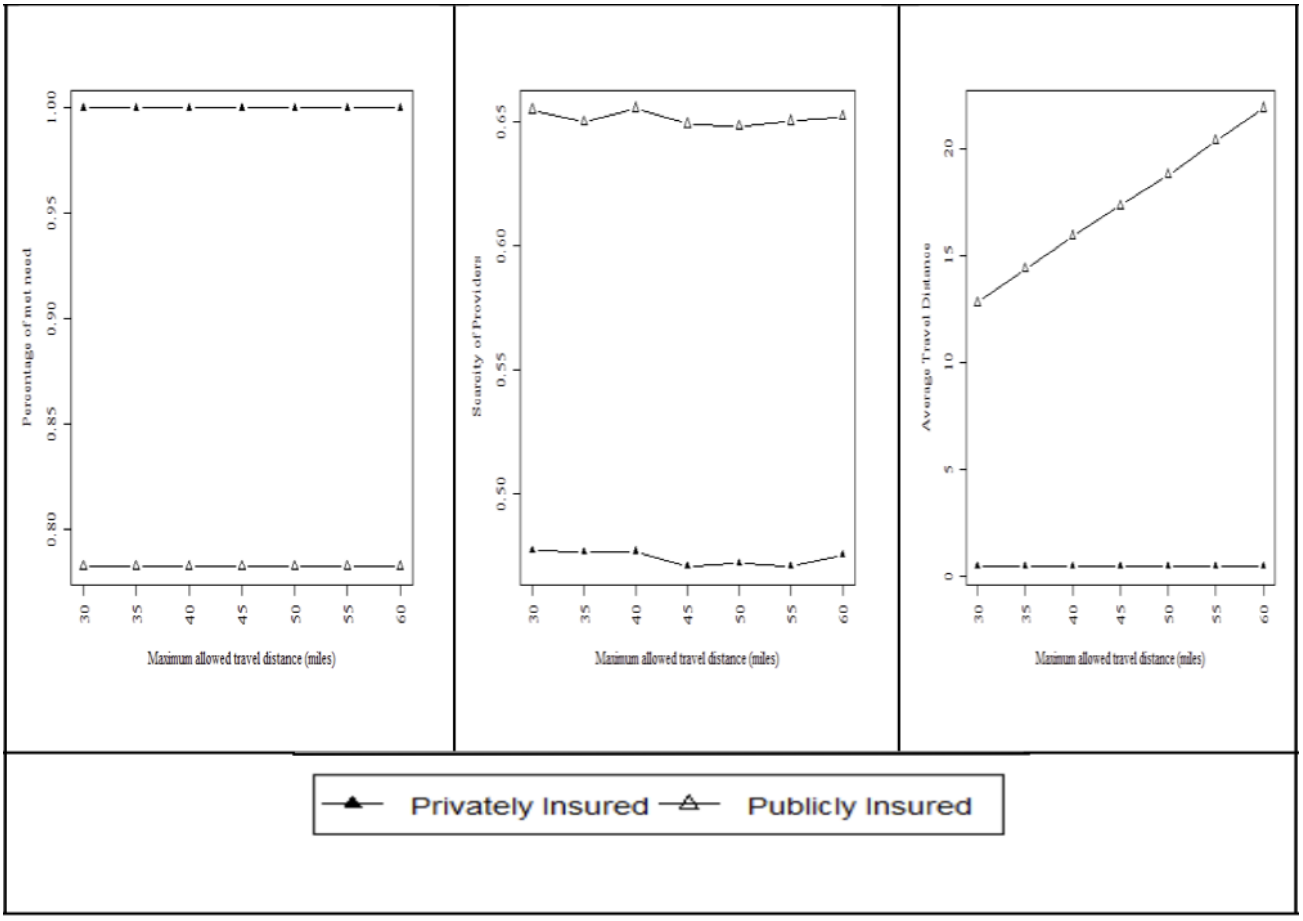

Figure 4.10: Sensitivity analysis of the percentage of met need, travel distance and scarcity of providers for urban areas with respect to changes in maximum allowed travel distance parameter.

4.4 Discussion

Approximately 60% of the 2.6 million children living in Georgia are eligible for public dental insurance. We found these children had significantly less access to dental care than privately insured children and disparities in access were most pronounced in rural areas. As a result, our model predicted that publicly insured children would travel at least 20 miles more to a dental office than would higher-income/privately insured children in 40% of Georgia census tracts, 50% of rural tracts and 35% of urban tracts.

Increasing the dentist participation rate from its current state level of 28% to 50% could

decrease the one-way travel distance for a dental visit from 40 to 25 miles and from 12 to 10 miles for publicly insured children living in rural and urban census tracts, respectively. The finding that almost doubling dentist participation in Medicaid would have negligible impact on privately insured children's access suggests that there is extra capacity that could become available if the Medicaid program in Georgia were to provide incentives to participate. One possible approach to accomplish this would be to raise Medicaid reimbursement rates. Although one study found that while increasing Medicaid reimbursements can be effective, the size of the effect can be modest [99, 100], a recent study found that increasing Medicaid reimbursement rates from roughly 35% of the private insurance reimbursement rate to 70% over a 4-year period while also simplifying administrative procedures during a recession (lowering private demand and thus opening dental capacity) in Connecticut increased utilization from 42% to 76% [101].

However, increasing dental fees at the level of incentivizing Medicaid participation may not be feasible in the current economic environment. Moreover, we find that only at a high Medicaid participation rate of 80%, all need for preventive dental care for publicly-insured children is met, which is very challenging to attain. Thus, another potential way to increase dental capacity for preventive dental care is to allow dental hygienists to provide preventive dental care in school settings [102]. This could also decrease costs as the marginal rate for a hygienist is significantly less than for a dentist. Given long travel distance in rural census tracts regardless of dentist participation rates, providing preventive dental care in schools might be an attractive solution [103]. Recently, Georgia passed legislation (HB-154) to allow dental hygienists to do this [104].

Our estimates are more conservative than those of the recently provided by the Health Policy Institute (HPI) of the American Dental Association[105]. According to HPI, 94% of children live within 15 minutes of a Medicaid dentists. Main reasons for the difference in the results lies in that our estimates account for the fact that dentists accepting Medicaid do not devote 100% of their capacity to Medicaid-enrolled children, assume that not all

dentists take new Medicaid-enrolled patients, and only focus on access to preventive care. Moreover, in recent research, optimization models as the one applied in this chapter have been compared to the classic catchment area method [106] used to provide the estimates by the HPI; while the optimization model is more complex, it has several advantages in terms of providing more accurate access estimates[107].

Limitations of this study pertain to assumptions made to estimate access and to the limited availability of detailed data. There are limitations in estimating need and supply for preventive dental care as highlighted in prior research [90]. First, we used household income thresholds for public insurance programs to estimate the numbers of children who are eligible for public insurance. Second, we relied on board of dentistry (BOD) data to identify practice location of providers. While many providers may practice from different offices, only the business address is provided in the BOD data. Third, we used the InsureKidsNow (IKN) database to identify providers accepting Medicaid, assuming capacity for public insurance to be within a given range. The IKN database has several inconsistencies, including repeat entries and many providers not found in the BOD data. We assumed all dentists in an office accepting public insurance took publicly-insured children. We also assumed that providers accepting public insurance took all types of public insurance. Fourth, we estimated matches between patients and providers assuming a centralized framework; in another work [96], we have shown how the model could be modified to incorporate decentralized decision making with patients maximizing their own individual welfare. Last, the travel distance does not account for potential differences is the associated travel time that may arise from population density or road shape.

The methodology in this study could help decision makers identify those areas where disparities in access to clinical preventive care are largest and implement strategies to increase dental capacity for Medicaid/CHIP patients, accordingly. Without access to preventive dental care it is likely that many of these children would develop dental caries. Dental caries is one of the most common diseases of childhood [108] and effective interventions

exist to prevent it [109, 110]. Further, there is evidence suggesting that increasing access to effective preventive dental services could be cost saving to CMS [111].

4.5 Acknowledgment

This chapter is part of the published papers in [90, 87].

Appendices

APPENDIX A

PROOFS IN CHAPTER 1

A.1 Properties of DC programming

The following are some known properties of the DC functions [31, 29].

1. Every DC function has a nonnegative DC decomposition; that is for a DC function f, there exists a decomposition, $f = g - h$, where both g and h are nonnegative and convex.

2. Every C^1 (i.e., functions with continuously first order derivatives) function with a Lipschitz gradient is a DC function.

3. Every twice continuously differentiable function is a DC function.

4. Every continuous function on a convex set is a limit of a sequence of uniformly converging DC functions.

5. Let f_i be DC functions for $i = 1, \cdots, m$. The DC functions are *closed* under the following operations:

   - summation: $\sum_{i=1}^{m} \lambda_i f_i(x)$, for $\lambda_i \in \mathbb{R}$, $i = 1, \cdots, m$
   - maximization: $\max_{i=1,\cdots,m} f_i(x)$
   - minimization: $\min_{i=1,\cdots,m} f_i(x)$
   - product: $\prod_{i=1,\cdots,m} f_i(x)$

6. A locally DC function that is defined in $\mathbb{R}^n$ is a DC function.

7. The following statements about a DC program are equivalent:

- $\sup\{f(x) : x \in C\}$, function f and set C are convex
- $\inf\{g(x) - h(x) : x \in \mathbb{R}^n\}$, functions g and h are convex
- $\inf\{g(x) - h(x) : x \in C, f_1(x) - f_1(x) \leq 0\}$, functions g, h, f_1, and f_2 and set C are all convex

Regarding the optimal solutions in the DC programming, the following have been developed in the literature [31, 29].

Definition A.1.1. (ϵ-subdifferential) *For a convex function $g(x)$, and $\epsilon > 0$, the ϵ-subdifferential of function $g(x)$ at point x_0 is denoted by $\partial_\epsilon g(x_0)$ and is defined as follows:*

$$\partial_\epsilon g(x_0) = \{\nu \in \mathbb{R}^n | g(x) \geq g(x_0) + \langle x - x_0, \nu \rangle - \epsilon\}.$$

One can verify that the subgradient [33, Chapter 23] (which is denoted by $\partial g(x_0)$) of function $g(x)$ at x_0 is the 0-subdifferential (i.e., $\epsilon = 0$).

Theorem A.1.1. (Global optimality condition) *A point x^* is a global optimal if and only if (iff) $\partial_\epsilon h(x^*) \subset \partial_\epsilon g(x^*)$ for any $\epsilon > 0$.*

Theorem A.1.2. (Local optimality condition) *A point x^* is a local optimal if $\partial h(x^*) \subset$ int $\partial g(x^*)$, where int $\partial g(x^*)$ represents the interior of the set $\partial g(x^*)$.*

A.2 Proofs in Section 1.4

A.2.1 Proof of Theorem 1.4.1

Proof. Since $\hat{\beta}_n^\lambda$ is d-stationary, the directional derivatives should be nonnegative in all directions, especially in the direction of $\beta^* - \hat{\beta}_n^\lambda$:

$$-\frac{1}{n}(\beta^* - \hat{\beta}_n^\lambda)^T X^T (Y - X\hat{\beta}_n^\lambda) + P'_\lambda(\hat{\beta}_n^\lambda; \beta^* - \hat{\beta}_n^\lambda) \geq 0,$$

where $P'_\lambda(\hat{\beta}_n^\lambda; \beta^* - \hat{\beta}_n^\lambda)$ is the directional derivative for the penalty function $P_\lambda = \lambda\|\beta\|_1 - h_\lambda(\beta)$ at $\hat{\beta}_n^\lambda$ in the direction of $\beta^* - \hat{\beta}_n^\lambda$.

According to Lemma 1.4.1, there exists a subgradient $z \in \partial g(\hat{\beta}_n^\lambda)$, where $\partial g(\hat{\beta}_n^\lambda)$ is the set of subgradient of $g(\beta) = \|\beta\|_1$ at $\hat{\beta}_n^\lambda$, such that:

$$\nabla L(\hat{\beta}_n^\lambda) + \lambda z - \nabla h_\lambda(\hat{\beta}_n^\lambda) = 0, \tag{A.2.1}$$

Multiplying by $(\beta^* - \hat{\beta}_n^\lambda)^T$ on both side and plugging in $Y = X\beta^* + \epsilon$, we have

$$-\frac{1}{n}(\beta^* - \hat{\beta}_n^\lambda)^T X^T X(\beta^* - \hat{\beta}_n^\lambda) - \frac{1}{n}(\beta^* - \hat{\beta}_n^\lambda)^T X^T \epsilon + P'_\lambda(\hat{\beta}_n^\lambda; \beta^* - \hat{\beta}_n^\lambda) = 0,$$

where without ambiguity, we let $P'_\lambda(\hat{\beta}_n^\lambda; \beta^* - \hat{\beta}_n^\lambda) = (\beta^* - \hat{\beta}_n^\lambda)^T(\lambda z - \nabla h_\lambda(\hat{\beta}_n^\lambda))$ since the true directional derivative for the penalty $P_\lambda = \lambda\|\beta\|_1 - h_\lambda(\beta)$ at $\hat{\beta}_n^\lambda$ in the direction of $\beta^* - \hat{\beta}_n^\lambda$ is greater than $(\beta^* - \hat{\beta}_n^\lambda)^T(\lambda z - \nabla h_\lambda(\hat{\beta}_n^\lambda))$. We thus will have

$$\frac{1}{n}(\beta^* - \hat{\beta}_n^\lambda)^T X^T X(\beta^* - \hat{\beta}_n^\lambda) = -\frac{1}{n}(\beta^* - \hat{\beta}_n^\lambda)^T X^T \epsilon + P'_\lambda(\hat{\beta}_n^\lambda; \beta^* - \hat{\beta}_n^\lambda), \tag{A.2.2}$$

which implies we have the following hold for $j \notin S$

$$\begin{aligned} & -\frac{1}{n}|(\hat{\beta}_n^\lambda)_j|\text{sign}((\hat{\beta}_n^\lambda)_j)X_j^T X(\beta^* - \hat{\beta}_n^\lambda) \\ =& \frac{1}{n}|(\hat{\beta}_n^\lambda)_j|\text{sign}((\hat{\beta}_n^\lambda)_j)X_j^T \epsilon - \lambda|(\hat{\beta}_n^\lambda)_j| + h'_\lambda((\hat{\beta}_n^\lambda)_j)|(\hat{\beta}_n^\lambda)_j|\text{sign}((\hat{\beta}_n^\lambda)_j) \\ \leq & -c\lambda|(\hat{\beta}_n^\lambda)_j|, \end{aligned} \tag{A.2.3}$$

where the "$\leq$" follows from Assumption 1.3.7.

For $j \in S$

$$
\begin{aligned}
&\frac{1}{n}(\beta^* - \hat{\beta}_n^\lambda)_j X_j^T X(\beta^* - \hat{\beta}_n^\lambda) \\
=&-\frac{1}{n}(\beta^* - \hat{\beta}_n^\lambda)_j X_j^T \epsilon + \lambda(\beta^* - \hat{\beta}_n^\lambda)_j z_j - h'_\lambda((\hat{\beta}_n^\lambda)_j)(\beta^* - \hat{\beta}_n^\lambda)_j \qquad \text{(A.2.4)} \\
\leq&\frac{5}{2}\lambda|(\beta^* - \hat{\beta}_n^\lambda)_j|,
\end{aligned}
$$

where the "$\leq$" follows from $\lambda \geq \frac{2\|X^T\epsilon\|_\infty}{n}$ and Assumption 1.3.2.

Let $\nu = \beta^* - \hat{\beta}_n^\lambda$, we thus will have

$$
\frac{1}{n}(\beta^* - \hat{\beta}_n^\lambda)^T X^T X(\beta^* - \hat{\beta}_n^\lambda) \leq -c\lambda\|(\nu)_{S^c}\|_1 + \frac{5}{2}\lambda\|(\nu)_S\|_1, \qquad \text{(A.2.5)}
$$

where the inequality follows from Assumption 1.3.2.

Since the left hand side of the above is nonnegative, we will have $\nu = \beta^* - \hat{\beta}_n^\lambda \in \mathcal{C}$. Under the restricted strong convexity condition, we will have

$$
\gamma\|\nu\|_2^2 \leq \frac{1}{n}(\beta^* - \hat{\beta}_n^\lambda)^T X^T X(\beta^* - \hat{\beta}_n^\lambda) \leq \frac{5}{2}\lambda\|(\nu)_S\|_1 \qquad \text{(A.2.6)}
$$

Thus we will further have

$$
\|\nu\|_2^2 \leq \frac{5}{2\gamma}\lambda\|(\nu)_S\|_1 \leq \frac{5}{2\gamma}\lambda\sqrt{|S|}\|(\nu)\|_2,
$$

from which we will have the upper bound

$$
\|\nu\|_2 \leq \frac{5}{2\gamma}\lambda\sqrt{|S|} \propto \frac{\lambda\sqrt{|S|}}{\gamma} \qquad \text{(A.2.7)}
$$

□

A.2.2 Proof of Corollary 1.4.1

Proof. From the proof of Theorem 1.4.1, let $\nu = \beta^* - \hat{\beta}_n^\lambda$,we have

$$\begin{aligned} &\|\frac{1}{\sqrt{n}}X(\beta^* - \hat{\beta}_n^\lambda)\|_2^2 \\ =&\frac{1}{n}(\beta^* - \hat{\beta}_n^\lambda)^T X^T X(\beta^* - \hat{\beta}_n^\lambda) \\ \leq& - c\lambda\|(\nu)_{S^c}\|_1 + \frac{5}{2}\lambda\|(\nu)_S\|_1 \\ \leq&(\frac{5}{2}\lambda)^2\frac{1}{\gamma}|S|. \end{aligned} \tag{A.2.8}$$

□

A.2.3 Proof of Corollary 1.4.2

Proof. Since ϵ_i for $i = 1, \cdots, n$ are from sub-Gaussian distribution with parameter σ^2,

$$\mathbb{P}\left(\frac{x_j^T\epsilon}{n} \geq t\right) \leq 2\exp\left(-\frac{nt^2}{2\sigma^2}\right)$$

By Bonferroni bound, we will have

$$\mathbb{P}(\frac{\|X^T\epsilon\|_\infty}{n} \geq t) \leq 2\exp\left(-\frac{nt^2}{2\sigma^2} + \log p\right)$$

By setting $t = \sigma\sqrt{\frac{\tau\log p}{n}}$ for some $\tau \geq 2$, we will be able to have

$$\mathbb{P}(\frac{\|X^T\epsilon\|_\infty}{n} \geq t) \leq 2\exp\left(-\frac{\tau-2}{2}\log p\right).$$

Thus:

$$\|\hat{\beta}_n^\lambda - \beta^*\|_2 \lesssim \frac{5}{\gamma}\sigma\sqrt{\frac{\tau|S|\log p}{n}} \tag{A.2.9}$$

□

A.2.4 Proof of Lemma 1.4.1

Proof. We will first prove the necessity. Since β_0 is a d-stationary solution to the objective function $F(\beta)$, we will have

$$F'(\beta_0, \beta - \beta_0) \geq 0,$$

for any $\beta \in \mathbb{R}^p$, where $F'(\beta_0, \beta - \beta_0)$ denotes the directional derivative in the direction of $\beta - \beta_0$. For any $i = 1, \cdots, p$, let $\beta^{i+} = \beta_0 + e_i$, where $e_i \in \mathbb{R}^p$ denotes the unit vector with 1 in the ith position and 0 everywhere else. Let $\beta^{i-} = \beta_0 - e_i$. We will have:

$$F'(\beta_0, \beta^{i+} - \beta_0) \geq 0,$$

$$F'(\beta_0, \beta^{i-} - \beta_0) \geq 0,$$

which implies

- For i such that $\beta_{0i} \neq 0$,

$$\nabla L(\beta_0)_i - \nabla h(\beta_0)_i + \text{sign}(\beta_{0i}) \geq 0,$$

$$-\nabla L(\beta_0)_i + \nabla h(\beta_0)_i - \text{sign}(\beta_{0i}) \geq 0,$$

  where $\text{sign}(x) = 1$ if $x > 0$, $\text{sign}(x) = -1$ if $x < 0$.

- For i such that $\beta_{0i} = 0$,

$$\nabla L(\beta_0)_i - \nabla h(\beta_0)_i + 1 \geq 0,$$

$$-\nabla L(\beta_0)_i + \nabla h(\beta_0)_i + 1 \geq 0.$$

We thus conclude that

- For i such that $\beta_{0i} \neq 0$,

$$\nabla L(\beta_0)_i - \nabla h(\beta_0)_i + \text{sign}(\beta_{0i}) = 0,$$

- For i such that $\beta_{0i} = 0$,

$$|\nabla L(\beta_0)_i - \nabla h(\beta_0)_i| \leq 1.$$

Thus for z such that $z_i = \text{sign}(\beta_{0i})$ when $\beta_{0i} \neq 0$ and $z_i = -(\nabla L(\beta_0)_i - \nabla h(\beta_0)_i)$ when $\beta_{0i} = 0$ is what we need.

On the other hand, if there exists some $z \in \partial g(\beta_0)$, where $\partial g(\beta_0)$ is the set of subgradient of $g(\beta)$ at β_0, such that:

$$\nabla L(\beta_0) + z - \nabla h(\beta_0) = 0, \tag{A.2.10}$$

- For i such that $\beta_{0i} \neq 0$,

$$z_i = \text{sign}(\beta_{0i}) = 1 \text{ or } -1,$$

- For i such that $\beta_{0i} = 0$,

$$-1 \leq z = -(\nabla L(\beta_0)_i - \nabla h(\beta_0)_i) \leq 1.$$

- For i such that $\beta_{0i} \neq 0$,

$$F'(\beta_0, \beta^{i+} - \beta_0) = 0,$$

$$F'(\beta_0, \beta^{i-} - \beta_0) = 0.$$

- For i such that $\beta_{0i} = 0$,

$$\begin{aligned} F'(\beta_0, \beta^{i+} - \beta_0) &= \nabla L(\beta_0)_i - \nabla h(\beta_0)_i + 1 \\ &= \nabla L(\beta_0)_i - \nabla h(\beta_0)_i + z + 1 - z \\ &= 0 + 1 - z \geq 0 \end{aligned} \tag{A.2.11}$$

$$\begin{aligned} F'(\beta_0, \beta^{i-} - \beta_0) &= -\nabla L(\beta_0)_i + \nabla h(\beta_0)_i + 1 \\ &= -\nabla L(\beta_0)_i + \nabla h(\beta_0)_i - z + 1 + z \\ &= 0 + 1 + z \geq 0 \end{aligned} \tag{A.2.12}$$

We thus conclude that the directional derivative of the objective function at β_0 is always nonnegative in any direction. This complete the proof. □

Remark A.2.1. *From the proof Lemma 1.4.1, we can derive similar conditions for "local maximals" for $\tilde{\beta}$ satisfying the following:*

$$F'(\tilde{\beta}, \beta - \tilde{\beta}) \leq 0. \tag{A.2.13}$$

- *For i such that $\tilde{\beta}_i \neq 0$,*

$$F'(\tilde{\beta}, \beta^{i+} - \tilde{\beta}) = 0,$$

$$F'(\tilde{\beta}, \beta^{i-} - \tilde{\beta}) = 0.$$

- *For i such that $\tilde{\beta}_i = 0$,*

$$\nabla L(\tilde{\beta})_i - \nabla h(\tilde{\beta})_i + 1 \leq 0,$$

$$-\nabla L(\tilde{\beta})_i + \nabla h(\tilde{\beta})_i + 1 \leq 0.$$

which implies that if the stationary solution $\tilde{\beta}$ to the FOC satisfies: $\min_{i=1}^{p}\{\tilde{\beta}_i\} = 0$, it will

only satisfy the condition for "local" minimals and thus be a d-stationary solution.

A.2.5 Proof of Lemma 1.4.2

Proof. Since $L(\beta) = \frac{1}{2n}\|Y - X\beta\|_2^2$ is quadratic and convex, we have

$$L(\beta_2) = L(\beta_1) + \nabla L(\beta_1)^T(\beta_2 - \beta_1) + \frac{1}{2}(\beta_2 - \beta_1)^T\nabla^2 L(\beta_1)(\beta_2 - \beta_1),$$

where $\nabla^2 L(\beta_1)$ is the Hessian matrix of $L(\beta)$ at β_1. Since $\nu = \beta_1 - \beta_2 \in \mathcal{C}$ and Assumption 1.3.8 holds on $\mathcal{C}$, we will further have

$$L(\beta_2) \geq L(\beta_1) + \nabla L(\beta_1)^T(\beta_2 - \beta_1) + \frac{\gamma}{2}\|\beta_2 - \beta_1\|_2^2.$$

On the other hand, $h_\lambda(\beta)$ is convex with $0 \leq \eta^+ \leq \frac{h'_\lambda(t_2) - h'_\lambda(t_1)}{t_2 - t_1} \leq \eta^-$, we will have

$$h_\lambda(\beta_2) \leq h_\lambda(\beta_1) + \nabla h_\lambda(\beta_1)^T(\beta_2 - \beta_1) + \frac{\eta^-}{2}\|\beta_2 - \beta_1\|_2^2$$

By combining the above two inequalities, we will be able to get (1.4.2). □

A.2.6 Proof of Theorem 1.4.2

Proof. The first part is easy to see since the feasible region is convex and is a subset of $\mathcal{C}$, in which the strong convexity condition holds (Assumption 1.3.8) for the loss function (in our case, the least square loss function). The minimizer to a strong problem is unique.
For the second conclusion, we first need to show that $X_S^T X_S$ is invertible and $\beta_S^O = (X_S^T X_S)^{-1} X_S^T Y$. This follows easily from the Assumption 1.3.8, which implies $\gamma_{\min}(X_S^T X_S)$, the minimum eigenvalue of $X_S^T X_S$ is larger than $n\gamma$. We thus have $\beta_S^O = (X_S^T X_S)^{-1} X_S^T Y$.

$$\begin{aligned} \beta_S^O - \beta_S^* &= (X_S^T X_S)^{-1} X_S^T Y - \beta^* \\ &= (X_S^T X_S)^{-1} X_S^T \epsilon \end{aligned} \tag{A.2.14}$$

Since $e_j(X_S^T X_S)^{-1} X_S^T \epsilon$, where $e_j \in \mathbb{R}^s$with all-zero elements except the j-th coordinate. Recall that ϵ has independent sub-Gaussian coordinates with the same variance parameter σ^2, we thus will have

$$\mathbb{P}\left(|e_j(X_S^T X_S)^{-1} X_S^T \epsilon| > t\right) \leq 2\exp - \frac{t^2}{\|e_j(X_S^T X_S)^{-1} X_S^T\|_2^2 \sigma^2}. \tag{A.2.15}$$

By using Bonferroni bound, the above implies

$$\mathbb{P}\left(\max_{j=1,\cdots,s} |e_j(X_S^T X_S)^{-1} X_S^T \epsilon| > t\right) \leq 2s\exp - \frac{t^2}{\|e_j(X_S^T X_S)^{-1} X_S^T\|_2^2 \sigma^2}. \tag{A.2.16}$$

Taking $t = C\|e_j(X_S^T X_S)^{-1} X_S^T \epsilon\|_2 \sigma \cdot \sqrt{2\log s}$ with $C > 0$, we will have

$$\begin{aligned} \|\beta_S^O - \beta_S^*\|_\infty &= \|(X_S^T X_S)^{-1} X_S^T \epsilon\|_\infty \\ &= \max_{j=1,\cdots,s} |e_j(X_S^T X_S)^{-1} X_S^T \epsilon| \\ &\leq C\|e_j(X_S^T X_S)^{-1} X_S^T \epsilon\|_2 \sigma \cdot \sqrt{2\log s} \end{aligned} \tag{A.2.17}$$

hold with probability at least $1 - 2\exp -C^2/s$. Since for any $j \in \{1, \cdots, s\}$,

$$\begin{aligned} \|e_j(X_S^T X_S)^{-1} X_S^T \epsilon\|_2^2 &= e_j(X_S^T X_S)^{-1} X_S^T \epsilon (e_j(X_S^T X_S)^{-1} X_S^T \epsilon^T) \\ &= e_j(X_S^T X_S)^{-1} e_j^T \\ &\leq 1/\gamma_{\min}(X_S^T X_S) \\ &\leq 1/n\gamma. \end{aligned} \tag{A.2.18}$$

This complete the proof since $\|\beta_{S^c}^O - \beta_{S^c}^*\|_\infty = 0$. □

A.2.7 Proof of Lemma 1.4.3

Proof. According to Theorem 1.4.2, we will have for $j \in S$,

$$|\beta_j^O| \geq |\beta_j^*| - \|\beta^O - \beta^*\|_\infty \geq \zeta,$$

which further implies $P_j'(\lambda, \beta_j^O) = 0$.

For $j \notin S$, we will have $h_\lambda'(\beta_j^O) = 0$ and since the errors are sub-Gaussian, there will exist $\xi_{S^c}^O \in \partial\|\beta_{S^c}^O\|_1$ satisfying inequality (1.4.4) with high probability and $\|\xi_{S^c}^O\|_\infty \leq \frac{1}{10}c$ where c is defined in the Assumption 1.3.7.

$$\left(\nabla L_n(\beta^O)\right)_{S^c} + \lambda\xi_{S^c}^O = 0.$$

In order to see this, first we notice that $\beta_S^O = (X_S^T X_S)^{-1} X_S^T Y$, $\beta_{S^c}^O = 0$. We thus will have $\left(\nabla L_n(\beta^O)\right)_{S^c} = -\frac{1}{n} X_{S^c}(Y - (X_S^T X_S)^{-1} X_S^T Y)$. Plugging in the true model $Y = X\beta^* + \epsilon$, we will have $\left(\nabla L_n(\beta^O)\right)_{S^c} = -\frac{1}{n} X_{S^c}(I - X_S(X_S^T X_S)^{-1} X_S^T)\epsilon$, where $(I - X_S(X_S^T X_S)^{-1} X_S^T)\epsilon$ is a vector of independent sub-Gaussian random variables. By using the Bonferroni bound, we will have the conclusion. □

A.2.8 Proof of Lemma 1.4.4

Proof. Since both $\hat{\beta}$ and β^O are d-stationary, per Lemma 1.4.3, we have:

$$(\hat{\beta} - \beta^O)^T \left(\nabla f_\lambda(\beta^O) + \lambda\xi^O\right) \geq 0,$$

and $\hat{\beta}$ is d-stationary, there exists a $\hat{\xi} \in \partial\|\hat{\beta}\|_1$ ((where $\partial\|\hat{\beta}\|_1$ stands for the subgradient of function $\|\beta\|_1$ at $\beta = \hat{\beta}$)) such that

$$(\beta^O - \hat{\beta})^T \left(\nabla f_\lambda(\hat{\beta}) + \lambda\hat{\xi}\right) \geq 0.$$

On the one hand,

$$\begin{aligned}
0 &\leq (\beta^O - \hat{\beta})^T \left(\nabla f_\lambda(\hat{\beta}) + \lambda \hat{\xi} \right) \\
&\leq (\beta^O - \hat{\beta})^T \left(\nabla f_\lambda(\hat{\beta}) \right) - \lambda \|(\nu)_{S^c}\|_1 + \lambda \|(\nu)_S\|_1 \\
&= -\frac{1}{n}(\beta^O - \hat{\beta})^T X^T X(\beta^* - \beta^O + \beta^O - \hat{\beta}) - \frac{1}{n}(\beta^O - \hat{\beta})^T X^T \epsilon \\
&- (\beta^O - \hat{\beta})^T \nabla h_\lambda(\hat{\beta}) - \lambda \|(\nu)_{S^c}\|_1 + \lambda \|(\nu)_S\|_1 \\
&\leq -\frac{1}{n}(\beta^O - \hat{\beta})^T X^T X(\beta^* - \beta^O + \beta^O - \hat{\beta}) - \frac{1}{n}(\beta^O - \hat{\beta})^T X^T \epsilon \\
&+ \sum_{i \notin S} |\hat{\beta}_i| |h'_\lambda(\hat{\beta}_i)| - \lambda \|(\nu)_{S^c}\|_1 + 2\lambda \|(\nu)_S\|_1.
\end{aligned} \tag{A.2.19}$$

By rearranging the terms, we will have

$$\begin{aligned}
&\frac{1}{n}(\beta^O - \hat{\beta})^T X^T X(\beta^* - \beta^O) + \frac{1}{n}(\beta^O - \hat{\beta})^T X^T X(\beta^O - \hat{\beta}) - \sum_{i \notin S} |\hat{\beta}_i| |h'_\lambda(\hat{\beta}_i)| \\
\leq &- \frac{1}{n}(\beta^O - \hat{\beta})^T X^T \epsilon - \lambda \|(\nu)_{S^c}\|_1 + 2\lambda \|(\nu)_S\|_1.
\end{aligned} \tag{A.2.20}$$

On the other hand, according to the proof of Lemma 1.4.3, we will have

$$\left(\nabla h_\lambda(\beta^O)) \right)_S = \lambda \text{sign}(\beta^O_S),$$

$$\left(\nabla h_\lambda(\beta^O)) \right)_{S^c} = 0,$$

$$\lambda \xi^O_{S^c} = - \left(\nabla L_n(\beta^O) \right)_{S^c},$$

$$\|\xi^O_{S^c}\|_\infty \leq \frac{1}{2}.$$

By using the above facts, we will further obtain

$$\begin{aligned} 0 &\leq (\hat{\beta} - \beta^O)^T \left(\nabla f_\lambda(\beta^O) + \lambda \xi^O\right) \\ &= -\frac{1}{n}(\hat{\beta} - \beta^O)^T X^T X(\beta^* - \beta^O) - \frac{1}{n}(\hat{\beta} - \beta^O)^T X^T \epsilon + (\hat{\beta} - \beta^O)^T_{S^c} \lambda \xi^O_{S^c} \\ &\leq -\frac{1}{n}(\hat{\beta} - \beta^O)^T X^T X(\beta^* - \beta^O) - \frac{1}{n}(\hat{\beta} - \beta^O)^T X^T \epsilon + \frac{1}{10} c\lambda \|(\nu)_{S^c}\|_1. \end{aligned} \tag{A.2.21}$$

By rearranging the terms, we will have

$$\frac{1}{n}(\hat{\beta} - \beta^O)^T X^T \epsilon - \frac{1}{10} c\lambda \|(\nu)_{S^c}\|_1 \leq \frac{1}{n}(\beta^O - \hat{\beta})^T X^T X(\beta^* - \beta^O) \tag{A.2.22}$$

Plugging inequality (A.2.22) to inequality (A.2.20), we will have

$$\frac{1}{n}(\beta^O - \hat{\beta})^T X^T X(\beta^O - \hat{\beta}) \leq \sum_{i \notin S} |\hat{\beta}_i| |h'_\lambda(\hat{\beta}_i)| - \lambda \|(\nu)_{S^c}\|_1 + \frac{1}{10} c\lambda \|(\nu)_{S^c}\|_1 + 2\lambda \|(\nu)_S\|_1. \tag{A.2.23}$$

Under the Assumption 1.3.7 with use of Bonferroni bound as in Corollary 1.4.2, we will have

$$0 \leq \frac{1}{n}(\beta^O - \hat{\beta})^T X^T X(\beta^O - \hat{\beta}) \leq -\frac{8}{10} c\lambda \|(\nu)_{S^c}\|_1 + 2\lambda \|(\nu)_S\|_1, \tag{A.2.24}$$

which implies $\lambda \|(\nu)_{S^c}\|_1 \leq \frac{5}{2c} \lambda \|(\nu)_S\|_1$ and $\nu \in \mathcal{C}$. □

A.2.9 Proof of Theorem 1.4.3

Proof. According to Lemma 1.4.2, we will have at $\hat{\beta}$ and β^O, respectively:

$$f_\lambda(\beta^O) \geq f_\lambda(\hat{\beta}) + \nabla f_\lambda(\hat{\beta})^T(\beta^O - \hat{\beta}) + \frac{\gamma - \eta^-}{2} \|\beta^O - \hat{\beta}\|_2^2,$$

$$f_\lambda(\hat{\beta}) \geq f_\lambda(\beta^O) + \nabla f_\lambda(\beta^O)^T(\hat{\beta} - \beta^O) + \frac{\gamma - \eta^-}{2} \|\hat{\beta} - \beta^O\|_2^2.$$

Since ℓ_1 norm penalty is convex, we will have

$$\lambda\|\hat{\beta}\|_1 \geq \lambda\|\beta^O\|_1 + \lambda(\hat{\beta} - \beta^O)^T\xi^O,$$

$$\lambda\|\beta^O\|_1 \geq \lambda\|\hat{\beta}\|_1 + \lambda(\beta^O - \hat{\beta})^T\hat{\xi},$$

where $\hat{\xi}$ and ξ^O are the same as in Lemma 1.4.4. Combine the above together, we will have:

$$0 \geq (\nabla f_\lambda(\hat{\beta}) + \lambda\hat{\xi})^T(\beta^O - \hat{\beta}) + (\nabla f_\lambda(\beta^O) + \lambda\xi^O)^T(\hat{\beta} - \beta^O) + (\gamma - \eta^-)\|\beta^O - \hat{\beta}\|_2^2.$$

Since both $\hat{\beta}$ and β^O are d-stationary, we will have

$$(\hat{\beta} - \beta^O)^T\left(\nabla f_\lambda(\beta^O) + \lambda\xi^O\right) \geq 0,$$

and $\hat{\beta}$ is d-stationary, there exists a $\hat{\xi} \in \nabla\{\|\hat{\beta}\|_1\}$ such that

$$(\beta^O - \hat{\beta})^T\left(\nabla f_\lambda(\hat{\beta}) + \lambda\hat{\xi}\right) \geq 0.$$

We thus will have $0 \geq (\gamma - \eta^-)\|\beta^O - \hat{\beta}\|_2^2$, which implies that $\beta^O = \hat{\beta}$. □

A.3 Proofs in Section 1.5

A.3.1 Proof of Lemma 1.5.1

Proof. Since $\hat{\beta}$ is a d-stationary solution to Problem (1.5.1), we have

$$\nabla L(\hat{\beta}) + \lambda z - \nabla h_\lambda(\hat{\beta}) = 0, \tag{A.3.1}$$

where $z \in \partial g(\hat{\beta})$, where $\partial g(\hat{\beta})$ is the set of subgradient of $g(\beta) = \|\beta\|_1$ at $\hat{\beta}$. We can get the gradient for the loss function

$$\nabla L(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}(\psi'(X_i^T\hat{\beta})X_i - Y_iX_i).$$

We can further write the above expression as

$$\nabla L(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\big((\psi'(X_i^T\hat{\beta})X_i - \psi'(X_i^T\beta^*)X_i) + (\psi'(X_i^T\beta^*)X_i - Y_iX_i)\big),$$

where the term $(\psi'(X_i^T\beta^*) - Y_iX_i)$ does not depend on the d-stationary solution. Multiply both side of (A.3.1) by $(\beta^* - \hat{\beta})^T$, we have

$$(\beta^*-\hat{\beta})^T\big\{\frac{1}{n}\sum_{i=1}^{n}\big((\psi'(X_i^T\hat{\beta})X_i-\psi'(X_i^T\beta^*)X_i)+(\psi'(X_i^T\beta^*)X_i-Y_iX_i)\big)+\lambda z-\nabla h_\lambda(\hat{\beta})\big\} = 0. \tag{A.3.2}$$

By rearranging the terms, we have

$$\begin{aligned}
0 &\le (\beta^* - \hat{\beta})^T\big\{\frac{1}{n}\sum_{i=1}^{n}(\psi'(X_i^T\beta^*)X_i - \psi'(X_i^T\hat{\beta})X_i)\big\} \\
&= (\beta^* - \hat{\beta})^T\big\{\frac{1}{n}\sum_{i=1}^{n}(\psi'(X_i^T\beta^*)X_i - Y_iX_i) + \lambda z - \nabla h_\lambda(\hat{\beta})\big\} \\
&\le (\beta^* - \hat{\beta})^T\big\{\frac{1}{n}\sum_{i=1}^{n}(\psi'(X_i^T\beta^*)X_i - Y_iX_i)\big\} + 2\lambda\|(\beta^* - \hat{\beta})_S\|_1 \\
&- \lambda\|\hat{\beta}_{S^c}\|_1 + \hat{\beta}_{S^c}^T\nabla h_\lambda(\hat{\beta}_{S^c}) \\
&\le (2 + \frac{c}{2})\lambda\|(\beta^* - \hat{\beta})_S\|_1 - \frac{c}{2}\lambda\|\hat{\beta}_{S^c}\|_1,
\end{aligned}$$

where the first "$\le$" is due to the convexity of the cumulant function, and the last one is due to the assumptions. We thus conclude that $\hat{\beta} \in \mathcal{C}$, which is defined in the Assumption 1.5.3. Given the restricted strong convexity, according to Lemma 1.4.2, let $f(\beta) = L(\beta) - h_\lambda(\beta)$,

we will have

$$f_\lambda(\beta^*) \geq f_\lambda(\hat{\beta}) + \nabla f_\lambda(\hat{\beta})^T(\beta^* - \hat{\beta}) + \frac{\gamma - \eta^-}{2}\|\beta^* - \hat{\beta}\|_2^2$$

and

$$f_\lambda(\hat{\beta}) \geq f_\lambda(\beta^*) + \nabla f_\lambda(\beta^*)^T(\hat{\beta} - \beta^*) + \frac{\gamma - \eta^-}{2}\|\beta^* - \hat{\beta}\|_2^2.$$

Adding the above up, we have

$$\nabla f_\lambda(\hat{\beta})^T(\hat{\beta} - \beta^*) \geq \nabla f_\lambda(\beta^*)^T(\hat{\beta} - \beta^*) + (\gamma - \eta^-)\|\beta^* - \hat{\beta}\|_2^2. \tag{A.3.3}$$

Adding $\lambda z^T(\hat{\beta} - \beta^*)$ to both side, we will have

$$0 = (\nabla f_\lambda(\hat{\beta})^T + \lambda z^T)(\hat{\beta} - \beta^*) \geq \nabla f_\lambda(\beta^*)^T(\hat{\beta} - \beta^*) + \lambda z^T(\hat{\beta} - \beta^*) + (\gamma - \eta^-)\|\beta^* - \hat{\beta}\|_2^2, \tag{A.3.4}$$

From inequalities (A.3.4) and (A.3.3), we have

$$\begin{aligned}(\gamma - \eta^-)\|\beta^* - \hat{\beta}\|_2^2 &\leq -\nabla f_\lambda(\beta^*)^T(\hat{\beta} - \beta^*) - \lambda z^T(\hat{\beta} - \beta^*) \\ &\leq (2 + \frac{c}{2})\lambda\|(\beta^* - \hat{\beta})_S\|_1 - \frac{c}{2}\lambda\|\hat{\beta}_{S^c}\|_1 \\ &\leq (2 + \frac{c}{2})\lambda\sqrt{|S|}\|\hat{\beta} - \beta^*\|_2\end{aligned} \tag{A.3.5}$$

where the last "$\leq$" is due to the fact that $\|(\hat{\beta} - \beta^*)_S\|_1 \leq \sqrt{|S|}\|\hat{\beta} - \beta^*\|_2$. We thus derive the bound that

$$\|\beta^* - \hat{\beta}\|_2 \leq \frac{(4 + c)\lambda}{2(\gamma - \eta^-)}\sqrt{|S|} \tag{A.3.6}$$

□

A.4 Proofs in Section 1.6

A.4.1 Proof of Lemma 1.6.1

Proof. Given β_k as the update in the kth iteration, we adopt the following procedure to update the estimation:

$$\beta_{k+1} = \arg\min L_n(\beta) + \sum_{i=1}^{p}(\lambda - h'(|\beta_{ki}|))|\beta_i|. \tag{A.4.1}$$

Let $Q(\beta|\beta_k) = L_n(\beta) + \lambda\|\beta_k\|_1 - h_\lambda(\beta_k) + \sum_{i=1}^{p}(\lambda - h'(|\beta_{ki}|))(|\beta_i| - |\beta_{ki}|)$. It can be easily checked that $Q(\beta_k|\beta_k) = F(\beta_k)$ and the following is equivalent to A.4.1:

$$\beta_{k+1} = \min Q(\beta|\beta_k). \tag{A.4.2}$$

Since $h_\lambda(\cdot)$ is convex by Assumption 1.3.4, we have

$$F(\beta) \leq Q(\beta|\beta_k).$$

According to the definition of β_{k+1}, we have

$$F(\beta_{k+1}) \leq Q(\beta_{k+1}|\beta_k) \leq Q(\beta_k|\beta_k) = F(\beta_k).$$

□

APPENDIX B

PROOFS IN CHAPTER 2

B.1 Proof of Lemma 2.3.1

Proof. Using Green's Identity, we have $|f|_{\Omega,m} \geq 0$ and

$$|f|_{\Omega,m} = \begin{cases} \int_\Omega |\Delta^{m/2} f(x)|^2 dx, & \text{even} \quad m, \\ \int_\Omega |\nabla(\Delta^{(m-1)/2} f(x))|^2 dx, & \text{odd} \quad m. \end{cases} \tag{B.1.1}$$

□

B.2 Proof of Proposition 2.3.1

Proof. Step 1: we define $\|y\| = \left(\min_{\phi \in H^m(\Omega), \phi(X_i)=y_i} |\phi|^2_{T,m}\right)^{1/2}$ and show that $\|\cdot\|$ is a semi-norm. To do so, we verify the following three properties:

1. Clearly, for any $y \in \mathbb{R}^n$, $\|y\| \geq 0$.

2. For any ϕ such that $\phi(X_i) = y_i$, since $\lambda\phi(X_i) = \lambda y_i$, we have $\|\lambda y\| \leq |\lambda\phi|_{T,m} = |\lambda||\phi|_{T,m}$. Thus, $\|\lambda y\| \leq |\lambda|\|y\|$.

3. Triangle inequality. Assume $y = y^1 + y^2$. For any $\varepsilon > 0$, there exist ϕ_1, ϕ_2, such that

$$\|y^1\| \geq |\phi_1|_{T,m} - \varepsilon, \quad \|y^2\| \geq |\phi_2|_{T,m} - \varepsilon.$$

   Then,

$$\|y^1\| + \|y^2\| \geq |\phi_1 + \phi_2|_{T,m} - 2\varepsilon.$$

Since we know that $(\phi_1 + \phi_2)(X_i) = y_i^1 + y_i^2 = y_i$ for $i = 1, 2, \cdots, n$, thus,

$$\|y^1\| + \|y^2\| \geq \|y\| - 2\varepsilon.$$

Let $\varepsilon \to 0$, we have $\|y^1\| + \|y^2\| \geq \|y\| = \|y^1 + y^2\|$.

Step 2: we show that $\|\cdot\|$ satisfies the following equality:

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2.$$

For $\forall \varepsilon > 0$, there exist ϕ_1, ϕ_2 s.t. $\phi_1(X_i) = u_i, \phi_2(X_i) = v_i, i = 1, 2, \cdots, n$ and $\|u\|^2 \geq |\phi_1|_{T,m}^2 - \varepsilon, \|v\|^2 \geq |\phi_2|_{T,m}^2 - \varepsilon$. Then,

$$\begin{aligned}
2\|u\|^2 + 2\|v\|^2 &\geq 2|\phi_1|_{T,m}^2 + 2|\phi_2|_{T,m}^2 - 4\varepsilon \\
&= 2\left|\frac{\phi_1 + \phi_2}{2} + \frac{\phi_1 - \phi_2}{2}\right|_{T,m}^2 + 2\left|\frac{\phi_1 + \phi_2}{2} - \frac{\phi_1 - \phi_2}{2}\right|_{T,m}^2 - 4\varepsilon \\
&= |\phi_1 + \phi_2|_{T,m}^2 + |\phi_1 - \phi_2|_{T,m}^2 - 4\varepsilon \\
&= \|u + v\|^2 + \|u - v\|^2 - 4\varepsilon.
\end{aligned}$$

Let $\varepsilon \to 0$, we have $\|u + v\|^2 + \|u - v\|^2 \leq 2\|u\|^2 + 2\|v\|^2$. Similarly, we replace u, v from above with $\frac{u+v}{2}$ and $\frac{u-v}{2}$, we get $\|u\|^2 + \|v\|^2 \leq \frac{1}{2}\|u + v\|^2 + \frac{1}{2}\|u - v\|^2$.

Step 3: we define a bilinear function based on $\|\cdot\|$:

$$\langle u, v\rangle \triangleq \frac{1}{4}\left(\|u + v\|^2 - \|u - v\|^2\right).$$

We only need to verify two properties, which are quite straightforward.

1. $\langle u_1 + u_2, v\rangle = \langle u_1, v\rangle + \langle u_2, v\rangle$.
2. $\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$.

Step 4: For any fixed v, define $\psi_v : \mathbb{R}^n \to \mathbb{R}$, $\psi_v(u) = \langle u, v\rangle$. Since $\langle u, v\rangle$ is bilinear,

ψ_v is linear, and therefore from Riesz representation theorem, there exists $w_v \in \mathbb{R}^n$, s.t. $\psi_v(u) = u^T w_v$.

Let $G : u \to w_v$. Clearly G is a linear mapping on $\mathbb{R}^n$. Hence there exists $E \in \mathbb{R}^{n\times n}$ s.t. $G(v) = Ev$ and $\langle u, v\rangle = u^T Ev$.

Since $\|u\|^2 = \langle u, u\rangle = u^T Eu$, E is semi-positive definite. Let $\mathbf{E}_{T,m} = nE$, we have

$$\frac{1}{n} y^T \mathbf{E}_{T,m} y = \langle y, y\rangle = \|y\|^2 = \min_{\phi\in H^m(\Omega),\phi(X_i)=y_i} |\phi|^2_{T,m}.$$

□

B.3 Proof of Lemma 2.3.2

Proof. Let V_d be the volume of unit ball in $\mathbb{R}^d$, and then $nV_d\delta^d_{\min} \le \mathrm{Vol}(\Omega)$. Therefore, we have

$$\delta^d_{\max} \le n^{-1}\frac{\mathrm{Vol}(\Omega)}{V_d}\frac{\delta^d_{\max}}{\delta^d_{\min}} = n^{-1}\frac{\mathrm{Vol}(\Omega)}{V_d}B^d_0 = O(n^{-1}),$$

from which we get $n\delta^d_{\max}$ is bounded from above.

Next, let u be a function satisfies

$$\frac{e_n}{n} = \frac{1}{n}\mathbf{u}^T\mathbf{E}_{T,1}\mathbf{u} = |u|_{T,1} = \min_{\phi\in H^1_\Omega,\phi(X_i)=u_i} |\phi|_{T,1},$$

where $\mathbf{u} = (u_1, \cdots, u_n)$ is the eigenvector of $\mathbf{E}_{T,1}$ corresponding to e_n. We define a compactly supported radial basis function

$$w(s) = \begin{cases} e^{-\|s\|/(1-\|s\|)}, & 0 \le \|s\| \le 1, \\ 0, & \|s\| > 1. \end{cases}$$

and specify $\phi(X) = \sum_{i=1}^n u_i w_i(X)$ where $w_i(X) = w(\frac{X-X_i}{\delta_{\min}})$. Clearly $\phi(X_i) = u_i$.

Moreover, we have for any multi-index $\alpha \in \mathbb{Z}_+^d$

$$D^\alpha w_i(X_j) = 0, \quad \forall i \neq j,$$

and with $|\alpha| = 2$

$$D^\alpha w_j(X_j) = \delta_{\min}^{-2} D^\alpha w(0).$$

Hence, we have

$$\begin{aligned}
|u|_{T,1}^2 &\leq |\phi|_{T,1}^2 \\
&= \frac{1}{n} \sum_{j=1}^n (\phi(X_j) \Delta \phi(X_j)) \\
&= \frac{1}{n} \sum_{j=1}^n \left(\sum_i u_i w_i(X_j) \right) \Delta \left(\sum_i u_i w_i(X_j) \right) \\
&= \frac{1}{n} \sum_{j=1}^n u_j^2 w_j(X_j) \Delta w_j(X_j) \\
&\leq \frac{1}{n} \sum_{j=1}^n u_j^2 \left(w_j(X_j) \sum_{|\alpha|=2} |D^\alpha w_j(X_j)| \right) \\
&\leq \frac{1}{n} \sum_{j=1}^n u_j^2 \left(\sum_{|\alpha|=2} |D^\alpha w(0)| \right) \delta_{\min}^{-2},
\end{aligned}$$

which implies that $e_n \leq C(w) \delta_{\min}^{-2}$, where $C(w) = \sum_{|\alpha|=2} |D^\alpha w(0)|$.

Finally we get

$$\delta_{\max}^2 e_n \leq C(w) \frac{\delta_{\max}^2}{\delta_{\min}^2} = C(w) B_0^2,$$

and prove that $\delta_{\max}^2 e_n$ is bounded from above. □

B.4 Auxiliary result B.4.1

Lemma B.4.1. *The function class $U_2^1(\Omega)$ has at least a subset as the polynomial spline space of degree $m+1$ with knots at $T = \{X_i\}_{i=1}^n$.*

Proof. Given $T = \{X_i\}_{i=1}^n$, we denote $\Pi : \{a = X_0 < X_1 < \cdots < X_n < X_{n+1} = b\}$ as the partition of $\Omega = [a, b]$. Let

$$\begin{aligned}\mathcal{S}_2(\Pi) = \{s(t) : s(t) = s_i(t) \in \mathcal{P}_2, t \in [X_i, X_{i+1}], i = 0, 1, \cdots, n;\\ D^l s_{i-1}(X_i) = D^l s_i(X_i), i = 1, \cdots, n, l = 0, 1; s'(a) = s'(b) = 0\}\end{aligned}$$

be a spline space of degree 2. The functional class $U_2^1(\Omega)$ is non-empty and it suffices to show that $\mathcal{S}_2(\Pi) \subset U_2^1(\Omega)$.

Obviously, $\mathcal{S}_2(\Pi) \subset \mathcal{W}_2^1(\Omega)$. For any $f \in \mathcal{S}_2(\Pi)$, assume $f(t)|_{[X_i, X_{i+1}]} = a_0 + a_1 t + a_2 t^2$, so we have

$$\begin{aligned}\frac{1}{2}[f'(X_i)^2 + f'(X_{i+1})^2] &= a_1^2 + 2a_1a_2(X_i + X_{i+1}) + 4a_2^2\frac{X_i^2 + X_{i+1}^2}{2},\\ \frac{1}{X_{i+1} - X_i}\int_{X_i}^{X_{i+1}} [f'(t)]^2 dt &= a_1^2 + 2a_1a_2(X_i + X_{i+1}) + 4a_2^2\frac{X_i^2 + X_iX_{i+1} + X_{i+1}^2}{3},\end{aligned}$$

From which we know $\frac{1}{2}[f'(X_i)^2 + f'(X_{i+1})^2] \geq \frac{1}{X_{i+1} - X_i}\int_{X_i}^{X_{i+1}} [f'(t)]^2 dt$. Besides,

$$\begin{aligned}&\frac{1}{2}[f'(X_i)^2 + f'(X_{i+1})^2]\\ \leq&\frac{1}{2}[f'(X_i)^2 + f'(X_{i+1})^2] + 2\left(a_1 + 2a_2\frac{X_i + X_{i+1}}{2}\right)^2\\ =&3[a_1^2 + 2a_1a_2(X_i + X_{i+1})] + 4a_2^2\left(\frac{X_i^2 + X_{i+1}^2}{2} + \frac{(X_i + X_{i+1})^2}{2}\right)\\ =&\frac{3}{X_{i+1} - X_i}\int_{X_i}^{X_{i+1}} [f'(t)]^2 dt.\end{aligned}$$

On one hand, we have

$$\begin{aligned}|f|_{T,1} &= \frac{1}{2n}\sum_{i=0}^{n}\left([f'(X_i)]^2 + [f'(X_{i+1})]^2\right) \\ &\geq \frac{1}{n}\sum_{i=0}^{n}\frac{1}{X_{i+1}-X_i}\int_{X_i}^{X_{i+1}}[f'(t)]^2dt \\ &\geq \frac{1}{n\delta_{\max}}|f|^2_{\Omega,1} \geq \frac{1}{B_0(b-a)}|f|^2_{\Omega,1}.\end{aligned}$$

On the other hand, we have

$$\begin{aligned}|f|_{T,1} &\leq \frac{1}{n}\sum_{i=0}^{n}\frac{3}{X_{i+1}-X_i}\int_{X_i}^{X_{i+1}}[f'(t)]^2dt \\ &\leq \frac{3}{n\delta_{\min}}|f|^2_{\Omega,1} \leq \frac{3B_0}{b-a}|f|^2_{\Omega,1}.\end{aligned}$$

Therefore, $\mathcal{S}_2(\Pi) \subset U_2^1(\Omega)$. □

B.5 Proof of Lemma 2.3.3

Proof. According to [112], there exists constant $C(d, m, \Omega, B_0) > 0$ and $\delta > 0$ such that for $\delta_{\max} \leq \delta_0$, we have

$$|f|^2_{T,0} \leq C(d, m, \Omega, B_0)\left(|f|^2_{\Omega,0} + \delta^2_{\max}|f|^2_{\Omega,1}\right).$$

Since $|f|^2_{T,1} \geq \underline{B}|f|^2_{\Omega,1}$, we have

$$\frac{|f|^2_{T,1}}{|f|^2_{T,0}} \geq \frac{\underline{B}|f|^2_{\Omega,1}}{C(d, m, \Omega, B_0)\left(|f|^2_{\Omega,0} + \delta^2_{\max}|f|^2_{\Omega,1}\right)} = \frac{|f|^2_{\Omega,1}}{C_1\left(|f|^2_{\Omega,0} + \delta^2_{\max}|f|^2_{\Omega,1}\right)},$$

where $C_1 = C(d, m, \Omega, B_0)/\underline{B}$. □

B.6 Proof of Lemma 2.3.4

Proof. According to [112], there exists constant $C(d, m, \Omega, B_0) > 0$ and $\delta_0 > 0$ such that for $\delta_{\max} \leq \delta_0$, we have

$$|f|^2_{\Omega,0} \leq C'(d, m, \Omega, B_0) \left(|f|^2_{T,0} + \delta^2_{\max} |f|^2_{T,1} \right).$$

Since $\underline{B}|f|^2_{\Omega,1} \leq |f|^2_{T,1} \leq \overline{B}|f|^2_{\Omega,1}$, we have

$$\frac{|f|^2_{\Omega,1}}{|f|^2_{\Omega,0}} \geq \frac{|f|^2_{T,1}/\overline{B}}{C'(d, m, \Omega, B_0) \left(|f|^2_{T,0} + \delta^2_{\max} |f|^2_{T,1}/\underline{B} \right)} \geq \frac{|f|^2_{T,1}}{C_2 \left(|f|^2_{T,0} + \delta^2_{\max} |f|^2_{T,1} \right)},$$

where $C_2 = \overline{B} C'(d, m, \Omega, B_0) \max(1, 1/\underline{B})$. □

B.7 Proof of Lemma 2.3.5

Proof. From Lemma 2.3.3, it holds

$$\frac{|\phi|^2_{T,1}}{|\phi|^2_{T,0}} \geq \frac{|\phi|^2_{\Omega,1}}{C_1 \left(|\phi|^2_{\Omega,0} + \delta^2_{\max} |\phi|^2_{\Omega,1} \right)}$$

for any $\phi \in H^m(\Omega)$ with $|\phi|^2_{T,0} \neq 0$. Then $e_j \geq \frac{1}{C_1}\theta_j$, where $\theta_1 \leq \cdots \leq \theta_n$ are the first n eigenvalues of the variational eigenvalue problem

$$|\phi|^2_{\Omega,1} = \theta \cdot \left(|\phi|^2_{\Omega,0} + \delta^2_{\max} |\phi|^2_{\Omega,1} \right),$$

which implies $\theta_j = \frac{\rho_j}{1+\delta^2_{\max}\rho_j}$, for any $j = 1, \cdots, n$.

Note that $\delta^2_{\max}\rho_j$ is bounded from above, since $\rho_j \sim j^{\frac{2}{d}}$ according to Theorem 14.6 in [62] and the fact that $\delta^2_{\max} = O(n^{-2/d})$ from Lemma 2.3.2. So there exists $C_3 > 0$ such that

$$\frac{1}{C_1(1 + \delta^2_{\max}\rho_j)} \geq C_3,$$

then we have $e_j \geq C_3 \rho_j$.

On the other hand, using Lemma 2.3.4, we have

$$\frac{|\phi|^2_{\Omega,1}}{|\phi|^2_{\Omega,0}} \geq \frac{|\phi|^2_{T,1}}{C_1\left(|\phi|^2_{T,0} + \delta^2_{\max}|\phi|^2_{T,1}\right)},$$

which implies $\rho_j \geq \frac{1}{C_2}\nu_j$, where $\nu_1 \leq \cdots \leq \nu_n$ are the first n eigenvalues of the variational eigenvalue problem

$$|\phi|^2_{T,1} = \nu \cdot \left(|\phi|^2_{T,0} + \delta^2_{\max}|\phi|^2_{T,1}\right),$$

which gives

$$\nu_j = \frac{e_j}{1 + \delta^2_{\max} e_j}, \quad j = 1, \cdots, n.$$

So there exists $C_4 > 0$ such that

$$e_j \leq C_2(1 + \delta^2_{\max} e_j)\rho_j \leq C_2(1 + \delta^2_{\max} e_n)\rho_j \leq C_4 \rho_j,$$

since $\delta^2_{\max} e_n$ is bounded according to the Lemma 2.3.2. □

B.8 Proof of Lemma 2.3.6

Proof. Let $d = \inf_{x \notin B} \|p - x\|^2$ and let $B^c = \Omega - B$. From B is open and Ω is locally compact, we know $d > 0$. Thus,

$$\left| \int_B e^{-\frac{\|p-y\|^2}{4t}} f(y)dy - \int_\Omega e^{-\frac{\|p-y\|^2}{4t}} f(y)dy \right| \leq \mu(B^c) \sup_{x \in B^c} (|f(x)|) e^{-\frac{d^2}{4t}},$$

where $\mu(B^c)$ denotes the Lebesgue measure of set B^c. The first two terms are constant and $e^{-\frac{d^2}{4t}}$ approaches 0 faster than any polynomial as $t \to 0$. □

B.9 Proof of Lemma 2.3.8

Proof. From Lemma 2.3.7, if we treat $f(p)$ as a constant function, we have

$$\frac{\partial}{\partial t}\left((4\pi t)^{-\frac{d}{2}}\int_{B(p)} e^{-\frac{\|p-y\|^2}{4t}} f(p)dy\right)|_{t=0} = Cf(p). \tag{B.9.1}$$

Denote $A(t) = (4\pi t)^{-\frac{d}{2}}\int_{\Omega} e^{-\frac{\|p-y\|^2}{4t}} f(y)dy$. Using the property of heat kernel, we have

$$A(0) = \lim_{t\to 0}(4\pi t)^{-\frac{d}{2}}\int_{B(p)} e^{-\frac{\|p-y\|^2}{4t}} f(p)dy = f(p). \tag{B.9.2}$$

Therefore,

$$\begin{aligned}\Delta f(p) &= \lim_{t\to 0}\frac{A(t)-A(0)}{t}\\ &= \lim_{t\to 0}(4\pi t)^{-\frac{d}{2}}\left(\int_{\Omega} e^{-\frac{\|p-y\|^2}{4t}} f(p)dy - \int_{\Omega} e^{-\frac{\|p-y\|^2}{4t}} f(y)dy\right).\end{aligned}$$

This completes our proof by setting $p = x_i, y = x_j$. □

B.10 Proof of Theorem 2.3.3

Proof. Without loss of generality, let $\mathbf{f}$ be the vector of function values at the knots of $T = \{X_i\}_{i=1}^n$ normalized by $|f|_{T,0}^2 = \frac{1}{n}\mathbf{f^T f} = \mathbf{1}$.

Then

$$\begin{aligned}
&\left|\frac{1}{nt^{d/2+1}}\frac{\mathbf{f^T L f}}{\mathbf{f^T f}} - \frac{|f|_{T,1}^2}{|f|_{T,0}^2}\right| \\
=&\left|\frac{1}{n^2t^{d/2+1}}\mathbf{f^T L f} - |\mathbf{f}|_{\mathbf{T},\mathbf{1}}^{\mathbf{2}}\right| \\
=&\left|\frac{1}{n^2t^{d/2+1}}\sum_{i=1}^{n} f(x_i)\sum_{j:j\neq i}(w_{i,j}f(x_i) - w_{i,j}f(x_j)) - \frac{1}{n}\sum_{i=1}^{n} f(x_i)\Delta f(x_i)\right| \\
\leq&\frac{M}{n}\sum_{i=1}^{n}\left|\frac{1}{nt^{d/2+1}}\sum_{j:j\neq i}(w_{i,j}f(x_i) - w_{i,j}f(x_j)) - \Delta f(x_i)\right|,
\end{aligned}$$

where M is the upper bound of $|f(X)|$ on Ω.

Define $Z = \frac{1}{n}\sum_{j:j\neq i}(w_{i,j}f_i - w_{i,j}f_j)$, and we know

$$\mathbb{E}[Z] = f(x_i)\int_\Omega w_{i,j}dx_j - \int_\Omega w_{i,j}f(x_j)dx_j.$$

Also from Hoeffding's Inequality,

$$\mathbb{P}\left(\frac{1}{t^{d/2+1}}|Z - \mathbb{E}[Z]| \geq \varepsilon\right) \leq e^{-\varepsilon^2 nt^{d+2}}.$$

Therefore,

$$\begin{aligned}
\left|\frac{1}{t^{d/2+1}}Z - \Delta f(x_i)\right| \leq& \underbrace{\left|t^{-\frac{d+2}{2}}(Z - \mathbb{E}[Z])\right|}_{(I)} \\
&+ \underbrace{\left|t^{-\frac{d+2}{2}}\left(\int_\Omega w_{i,j}f(x_i)dx_j - \int_\Omega w_{i,j}f(x_j)dx_j\right) - \Delta f(x_i)\right|}_{(II)}.
\end{aligned}$$

Since

$$(I) \leq e^{-\varepsilon^2 nt^{d+2}}t^{-\frac{d+2}{2}}|Z - \mathbb{E}[Z]| + (1 - e^{-\varepsilon^2 nt^{d+2}})\varepsilon \to 0,$$

as $n \to \infty$ if $t_n = O(n^{-\frac{1}{d+2+\alpha}})$ where $\alpha > 0$.

Besides, from Lemma 2.3.8, we know $(II) \to 0$ as $n \to \infty$. Therefore all the above

implies $|\mu_j - e_j| \to 0$ as $n \to \infty$. Hence from Theorem 2.3.2, we have

$$C_7 j^{2/d} \leq \mu_j \leq C_8 j^{2/d}.$$

For the case $m > 1$, we have the following clearly

$$C_7 j^{2m/d} \leq \mu_j \leq C_8 j^{2m/d}.$$

Then our proof is completed. □

B.11 Proof of Theorem 2.3.5

We will need the following lemma.

Lemma B.11.1. *If for $B_3, B_4 > 0$, we have $B_3 j^m \leq \mu_j \leq B_4 j^m$ for a constant $m > 0$ and $j = 1, 2, \cdots$ then the following holds for $n > 0, \lambda > 0$,*

$$\sum_{j=1}^{n} \frac{1}{(1 + \lambda \mu_j)^2} = O(\lambda^{-\frac{1}{m}}). \tag{B.11.1}$$

Proof. Evidently we have

$$\sum_{j=1}^{m} \frac{1}{(1 + \lambda B_2 j^m)^2} \leq \sum_{j=1}^{n} \frac{1}{(1 + \lambda \mu_j)^2} \leq \sum_{j=1}^{n} \frac{1}{(1 + \lambda B_1 j^m)^2}.$$

For $i = 1, 2$, we have

$$\begin{aligned}
\sum_{j=1}^{n} \frac{1}{(1+\lambda B_i j^m)^2} &\geq \int_1^{n+1} \frac{1}{(1+\lambda B_i x^m)^2} dx \\
&= \frac{1}{m} \int_{\lambda B_i}^{\lambda B_i (n+1)^m} \frac{y^{-\frac{m-1}{m}}}{(1+y)^2} dy \cdot (\lambda B_i)^{-\frac{1}{m}} \\
&\to m^{-1} \left(\int_{\lambda B_i}^{\infty} \frac{y^{-\frac{m-1}{m}}}{(1+y)^2} dy \right) B_i^{-1/m} \cdot \lambda^{-1/m} \\
&= O(\lambda^{-1/m}),
\end{aligned}$$

where the second equation comes from the change of variable $y = \lambda B_i x^m$. Similarly we also have the following

$$\begin{aligned}
\sum_{j=1}^{n} \frac{1}{(1+\lambda B_i j^m)^2} &\leq \int_0^{n} \frac{1}{(1+\lambda B_i x^m)^2} dx \\
&= \frac{1}{m} \int_{0}^{\lambda B_i n^m} \frac{y^{-\frac{m-1}{m}}}{(1+y)^2} dy \cdot (\lambda B_i)^{-\frac{1}{m}} \\
&\to m^{-1} \left(\int_{\lambda B_i}^{\infty} \frac{y^{-\frac{m-1}{m}}}{(1+y)^2} dy \right) B_i^{-1/m} \cdot \lambda^{-1/m} \\
&= O(\lambda^{-1/m}).
\end{aligned}$$

□

We now give the proof of Theorem 2.3.5.

Proof. By using the bounds of eigenvalues of matrix $\mathbf{M}$ that $\mu_j = O(j^{\frac{2m}{d}})$ obtained from

Theorem 2.3.3, we have

$$\begin{aligned}
\mathbb{E}[r_n(\lambda)] &= \mathbb{E}[n^{-1}\|\hat{\mathbf{f}}_n(\lambda) - \mathbf{f}\|^{\mathbf{2}}] \\
&= n^{-1}\left(\mathbf{f^T}(\mathbf{A_n}(\lambda) - \mathbf{I})^{\mathbf{2}}\mathbf{f} + \sigma^{\mathbf{2}}\mathrm{tr}[\mathbf{A_n}(\lambda)^{\mathbf{2}}]\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}\frac{\lambda^2\mu_j^2 b_j^2}{(1+\lambda\mu_j)^2} + \frac{\sigma^2}{n}\sum_{j=1}^{n}\frac{1}{(1+\lambda\mu_j)^2} \\
&\leq \frac{\lambda}{n}\sum_{j=1}^{n}\frac{\lambda\mu_j}{(1+\lambda\mu_j)^2}\mu_j b_j^2 + \frac{\sigma^2}{n}\sum_{j=1}^{n}\frac{1}{(1+\lambda\mu_j)^2} \\
&\leq \frac{\lambda}{4n}\mathbf{f^T M f} + \frac{\sigma^{\mathbf{2}}}{\mathbf{n}}\sum_{\mathbf{j=1}}^{\mathbf{n}}\frac{\mathbf{1}}{(\mathbf{1}+\lambda\mu_{\mathbf{j}})^{\mathbf{2}}} \\
&= O(\lambda) + O(n^{-1}\lambda^{-\frac{d}{2m}}),
\end{aligned}$$

where $\mathbf{b} = \mathbf{U}^T\mathbf{f} = (\mathbf{b_1}, \mathbf{b_2}, \cdots, \mathbf{b_n})^{\mathbf{T}}$, $\mathbf{M} = \mathbf{U}\Lambda\mathbf{U}^T$ and the last "=" comes from the Lemma B.11.1. □

B.12 Proof of Lemma 2.3.9

Proof. Let $0 = \mu_1 < \mu_2 \leq \cdots \leq \mu_n$ be the eigenvalues of penalty matrix $\mathbf{M}$, and u_j the unit eigenvector corresponding to μ_j, $j = 1, 2, \cdots, n$. So we have

$$\begin{aligned}
n\mathbb{E}[r_n(\lambda)] &= n\mathbb{E}[n^{-1}\|\hat{\mathbf{f}}_n(\lambda) - \mathbf{f}\|^{\mathbf{2}}] \\
&= E[(\hat{\mathbf{f}}_n(\lambda) - \mathbf{f})^{\mathbf{T}}(\hat{\mathbf{f}}_{\mathbf{n}}(\lambda) - \mathbf{f})] \\
&= \mathbf{f^T}(\mathbf{A_n}(\lambda) - \mathbf{I})^{\mathbf{2}}\mathbf{f} + \sigma^{\mathbf{2}}\mathrm{tr}[\mathbf{A_n}(\lambda)^{\mathbf{2}}] \\
&= \sum_{j=2}^{n}\frac{\lambda^2\mu_j^2}{(1+\lambda\mu_j)^2}b_j^2 + \sigma^2\sum_{j=1}^{n}\frac{1}{(1+\lambda\mu_j)^2},
\end{aligned}$$

where $b_j = \mathbf{u_j^T f}$.

If $\lambda \sim O(1)$ or $\lambda \to \infty$, since $\mu_j \sim j^{2m/d}$ for $j > 1$, there exists j^* such that $j^*/n \to 0$ and $\frac{\lambda\mu_j}{1+\lambda\mu_j} \geq \frac{1}{2}$ for $j > j^*$, then

$$n\mathbb{E}[r_n(\lambda)] \geq \sum_{j=2}^{n} \frac{\lambda^2\mu_j^2}{(1+\lambda\mu_j)^2} b_j^2 \geq \frac{1}{4}\sum_{j>j^*} b_j^2$$
$$\geq \frac{n}{4}|f|_{T,0}^2 - \frac{1}{4}j^* \max\{b_1^2, \cdots, b_{j^*}^2\} = O(n) \to \infty.$$

On the other hand, if $\lambda \to 0$ as $n \to \infty$, we have

$$n\mathbb{E}[r_n(\lambda)] = \sigma^2 \sum_{j=1}^{n} \frac{1}{(1+\lambda\mu_j)^2} = O(\lambda^{-\frac{d}{2m}}) \to \infty,$$

where the second equation is based on Lemma B.11.1. □

B.13 Proof of Lemma 2.3.10

Due to the following decomposition of risk function:

$$\begin{aligned}
&|\mathbb{E}[r_n(\lambda)] - r_n(\lambda)| \\
=&n^{-1}\left|\mathbf{f}^T(\mathbf{A}_n(\lambda)-\mathbf{I})^2\mathbf{f} + \sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2] - \|(\mathbf{A}_n(\lambda)-\mathbf{I})\mathbf{f} + \mathbf{A}_n(\lambda)\varepsilon\|\right| \\
=&n^{-1}\left|\|\mathbf{A}_n(\lambda)\varepsilon\|^2 - \sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2] + 2\mathbf{f}^T\left(\mathbf{A}_n(\lambda)-\mathbf{I}_n\right)\mathbf{A}_n(\lambda)\varepsilon\right|.
\end{aligned}$$

it suffices to show

$$\sup_{\lambda>0} \frac{n^{-1}|\mathbf{f}^T\left(\mathbf{A}_n(\lambda)-\mathbf{I}_n\right)\mathbf{A}_n(\lambda)\varepsilon|}{\mathbb{E}[r_n(\lambda)]} \xrightarrow{p} 0, \tag{B.13.1}$$

and

$$\sup_{\lambda>0} \frac{n^{-1}\left|\|\mathbf{A}_n(\lambda)\varepsilon\|^2 - \sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2]\right|}{\mathbb{E}[r_n(\lambda)]} \xrightarrow{p} 0. \tag{B.13.2}$$

According to *Chebyshev Inequality*, for any given $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\frac{n^{-1}|\mathbf{f}^T\left(\mathbf{A}_n(\lambda)-\mathbf{I}_n\right)\mathbf{A}_n(\lambda)\varepsilon|}{\mathbb{E}[r_n(\lambda)]} > \delta\right)\\
\leq&\delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-2}\mathbb{E}[(\mathbf{f}^T\left(\mathbf{A}_n(\lambda)-\mathbf{I}_n\right)\mathbf{A}_n(\lambda)\varepsilon)^2]\\
=&\delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-2}\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)\left(\mathbf{A}_n(\lambda)-\mathbf{I}_n\right)\mathbf{f}\mathbf{f}^T\left(\mathbf{A}_n(\lambda)-\mathbf{I}_n\right)\mathbf{A}_n(\lambda)]\\
=&\delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-2}\sigma^2\|\mathbf{A}_n(\lambda)\left(\mathbf{A}_n(\lambda)-\mathbf{I}_n\right)\mathbf{f}\|^2\\
\leq&\delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-1}\sigma^2\frac{\|\left(\mathbf{A}_n(\lambda)-\mathbf{I}_n\right)\mathbf{f}\|^2}{n\mathbb{E}[r_n(\lambda)]}\\
\leq&\sigma^2(n\mathbb{E}[r_n(\lambda)])^{-1}\to 0,
\end{aligned}$$

where the second to last "$\leq$" is due to the eigen-decomposition of $\mathbf{A}_n(\lambda) = \mathbf{U}\Lambda\mathbf{U}^T$ and the diagonal elements of Λ are less than 1, and the last "$\leq$" is due to $n\mathbb{E}[r_n(\lambda)] \geq \|\left(\mathbf{A}_n(\lambda)-\mathbf{I}_n\right)\mathbf{f}\|^2$. Thus (B.13.1) holds.

Again for any given $\delta > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{n^{-1}\left|\|\mathbf{A}_n(\lambda)\varepsilon\|^2-\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2]\right|}{\mathbb{E}[r_n(\lambda)]} > \delta\right)\\
\leq&\delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-2}\mathbb{E}[\left(\|\mathbf{A}_n(\lambda)\varepsilon\|^2-\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2]\right)^2]\\
=&\delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-1}\frac{E[\|\mathbf{A}_n(\lambda)\varepsilon\|^4]-(\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2])^2}{n\mathbb{E}[r_n(\lambda)]}.
\end{aligned}$$

The last equality comes from the fact that $\mathbb{E}(\|\mathbf{A}_n(\lambda)\varepsilon\|^2) = \sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2]$ and

$$\mathbb{E}[\left(\|\mathbf{A}_n(\lambda)\varepsilon\|^2-\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2]\right)^2] = \mathrm{Var}(\|\mathbf{A}_n(\lambda)\varepsilon\|^2).$$

We know from the beginning that $n\mathbb{E}[r_n(\lambda)] \geq \sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2]$, it suffices to show that

$$\frac{E[\|\mathbf{A}_n(\lambda)\varepsilon\|^4]-(\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2])^2}{\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2]} < \text{Constant}. \tag{B.13.3}$$

Denote $\mathbf{B} = \mathbf{A}_n(\lambda)^2 = (B_{ij})_{n\times n}$, then we have

$$\begin{aligned}
\mathbb{E}[\|\mathbf{A}_n(\lambda)\varepsilon\|^4] &= \mathbb{E}[(\varepsilon^T\mathbf{B}\varepsilon)^2] \\
&= \mathbb{E}[(\sum_{i,j} B_{ij}\varepsilon_i\varepsilon_j)(\sum_{k,l} B_{kl}\varepsilon_k\varepsilon_l)] \\
&\leq \left(\sum_{i=1}^n B_{ii}\sigma^2\right)^2 + \sum_{i=1}^n B_{ii}^2 E[\varepsilon_i^4] + \sum_{i\neq j} B_{ij}^2\sigma^4.
\end{aligned}$$

There exists a constant c such that $\mathbb{E}[\varepsilon_i^4] \leq c\sigma^2$ and $\sigma^4 \leq c\sigma^2$, so we get

$$\begin{aligned}
\mathbb{E}[\|\mathbf{A}_n(\lambda)\varepsilon\|^4] &= \left(\sum_{i=1}^n B_{ii}\sigma^2\right)^2 + c\sum_{i=1}^n B_{ii}^2\sigma^2 + c\sum_{i\neq j} B_{ij}^2\sigma^2 \\
&= \left(\sum_{i=1}^n B_{ii}\sigma^2\right)^2 + c\sum_{i,j} B_{ij}^2\sigma^2 \\
&= (\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2])^2 + c\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^4] \\
&\leq (\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2])^2 + c\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2],
\end{aligned}$$

where the last "$\leq$" is due to $\mathbf{A}_n(\lambda) = \mathbf{U}\Lambda\mathbf{U}^T$ and the the diagonal elements of Λ are less than 1. This completes (B.13.3) and completes the proof.

B.14 Proof of Lemma 2.3.11

From Section. 2.3.2 we know $\mu_i = O(i^{\frac{2m}{d}})$. Then

$$\begin{aligned}
\lim_{n\to\infty} \frac{\left(n^{-1}\sum_{i=l+1}^{n} k_i^{-1}\right)^2}{n^{-1}\sum_{i=l+1}^{n} k_i^{-2}} &= \lim_{n\to\infty} \frac{\left(\sum_{i=l+1}^{n} \mu_i^{-1}\right)^2}{n\sum_{i=l+1}^{n} \mu_i^{-2}} \\
&= \lim_{n\to\infty} \frac{\left(\int_{l+1}^{n} \mu^{-2m/d} d\mu\right)^2}{n\int_{l+1}^{n} \mu^{-4m/d} d\mu} \\
&= \lim_{n\to\infty} \frac{(4m-d)d}{(2m-d)^2} \cdot \frac{\left((l+1)^{1-\frac{2m}{d}} - n^{1-\frac{2m}{d}}\right)^2}{n\left((l+1)^{1-\frac{4m}{d}} - n^{1-\frac{4m}{d}}\right)} \\
&= \lim_{n\to\infty} \frac{(4m-d)d}{(2m-d)^2} \cdot \frac{l+1}{n} \cdot \frac{\left(1-(\frac{l+1}{n})^{\frac{2m}{d}-1}\right)^2}{1-(\frac{l+1}{n})^{\frac{4m}{d}-1}} \\
&= 0.
\end{aligned}$$

B.15 Proof of Lemma 2.3.12

Proof. From the fact that

$$\sigma^2(n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda_n)])^2 \leq \sigma^2 n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda_n)^2] \leq \mathbb{E}[r_n(\lambda_n)] \to 0,$$

therefore $n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda_n)] \to 0$, and thus

$$n^{-1}\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\lambda_n)] \to 1.$$

By the fact that $n^{-1}\|\varepsilon\|^2 \to \sigma^2$ and Cauchy-Schwartz inequality,

$$n^{-1}\|(\mathbf{I}_n - \mathbf{A}_n(\lambda_n))\mathbf{y}\|^2 = n^{-1}\|\varepsilon\|^2 + n^{-1}\|\mathbf{f} - \hat{\mathbf{f}}_n(\lambda_n)\|^2 + \frac{2}{n}\left|(\mathbf{f} - \hat{\mathbf{f}}_n(\lambda_n))^T\varepsilon\right| \to \sigma^2.$$

□

B.16 Proof of Lemma 2.3.13

Proof. Recall $\mathbf{A}_n(\lambda_n) = (\mathbf{I}_n + \lambda_n \mathbf{M})^{-1} = (\mathbf{I}_n + \mathbf{K}_n(\lambda_n))^{-1}$. Obviously

$$\frac{(n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda_n)])^2}{n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda_n)^2]} = \frac{(n^{-1}\sum_{i=1}^n (1+\kappa_i)^{-1})^2}{n^{-1}\sum_{i=1}^n (1+\kappa_i)^{-2}}, \tag{B.16.1}$$

where $0 \leq \kappa_1 \leq \cdots \leq \kappa_n$ are the eigenvalues of $\mathbf{K}_n(\lambda_n)$. Let l be the number holding $\kappa_l \leq 1 < \kappa_{l+1}$, then we have

$$\sum_{i=1}^n (1+\kappa_i)^{-1} \leq l + \sum_{i=l+1}^n \kappa_i^{-1},$$

and

$$\sum_{i=1}^n (1+\kappa_i)^{-2} \geq \frac{1}{4}\left(l + \sum_{i=l+1}^n \kappa_i^{-2}\right).$$

Then it suffices to show

$$\left(\frac{l}{n} + \frac{1}{n}\sum_{i=l+1}^n \kappa_i^{-1}\right)^2 / \frac{1}{4}\left(\frac{l}{n} + \frac{1}{n}\sum_{i=l+1}^n \kappa_i^{-2}\right) \to 0.$$

$\mathbb{E}[r_n(\lambda_n)] \to 0$, since $r_n(\lambda_n)$ is nonnegative, thus we have $n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda_n)^2] \to 0$, from which we arrive at our result. □

B.17 Proof of Lemma 2.3.14

Proof. We first prove (2.3.16), which can be rewritten as

$$\begin{aligned}
&\frac{2\left|\frac{\sigma^2\mathrm{tr}[\mathbf{I}_n-\mathbf{A}_n(\lambda)]\mathbf{y}^T(\mathbf{I}_n-\mathbf{A}_n(\lambda))\varepsilon}{n\|(\mathbf{I}_n-\mathbf{A}_n(\lambda))\mathbf{y}\|^2}-\frac{\sigma^4(\mathrm{tr}[\mathbf{I}_n-\mathbf{A}_n(\lambda)])^2}{n\|(\mathbf{I}_n-\mathbf{A}_n(\lambda))\mathbf{y}\|^2}-n^{-1}\|\varepsilon\|^2+\sigma^2\right|}{r_n(\lambda)}\\
\leq& 2\frac{\sigma^2\mathrm{tr}[\mathbf{I}_n-\mathbf{A}_n(\lambda)]}{\|(\mathbf{I}_n-\mathbf{A}_n(\lambda))\mathbf{y}\|^2}\cdot\frac{n^{-1}\left|\mathbf{f}^T(\mathbf{I}_n-\mathbf{A}_n(\lambda))\varepsilon\right|}{r_n(\lambda)}\\
&+2\frac{\sigma^2\mathrm{tr}[\mathbf{I}_n-\mathbf{A}_n(\lambda)]}{\|(\mathbf{I}_n-\mathbf{A}_n(\lambda))\mathbf{y}\|^2}\cdot\frac{n^{-1}\left|\varepsilon^T\mathbf{A}_n(\lambda)\varepsilon-\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)]\right|}{r_n(\lambda)}\\
&+2\frac{\left|\left(\frac{\sigma^2\mathrm{tr}[\mathbf{I}_n-\mathbf{A}_n(\lambda)]}{\|(\mathbf{I}_n-\mathbf{A}_n(\lambda))\mathbf{y}\|^2}-1\right)(\sigma^2-n^{-1}\|\varepsilon\|^2)\right|}{r_n(\lambda)}. \qquad \text{(B.17.1)}
\end{aligned}$$

Note that $n^{-1}\mathrm{tr}[\mathbf{I}_n-\mathbf{A}_n(\lambda_n)]\to 1$, $n^{-1}\|(\mathbf{I}_n-\mathbf{A}_n(\lambda_n))\mathbf{y}\|^2\to\sigma^2$, from Lemma 2.3.12 and Lemma 2.3.10, it suffices to show the following three equations

$$\sup_{\lambda>0}\frac{n^{-1}\left|\mathbf{f}^T(\mathbf{I}_n-\mathbf{A}_n(\lambda))\varepsilon\right|}{\mathbb{E}[r_n(\lambda)]}\to 0, \qquad \text{(B.17.2)}$$

$$\sup_{\lambda>0}\frac{n^{-1}\left|\varepsilon^T\mathbf{A}_n(\lambda)\varepsilon-\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)]\right|}{\mathbb{E}[r_n(\lambda)]}\to 0, \qquad \text{(B.17.3)}$$

$$\sup_{\lambda>0}\frac{|(\sigma^2n^{-1}\mathrm{tr}[\mathbf{I}_n-\mathbf{A}_n(\lambda)]-n^{-1}\|(\mathbf{I}_n-\mathbf{A}_n(\lambda))\mathbf{y}\|^2)(\sigma^2-n^{-1}\|\varepsilon\|^2)|}{\mathbb{E}[r_n(\lambda)]}\to 0. \qquad \text{(B.17.4)}$$

For (B.17.2), according to Chebyshev Inequality, for any given $\delta > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{n^{-1}\left|\mathbf{f}^T(\mathbf{I}_n - \mathbf{A}_n(\lambda)\varepsilon)\right|}{\mathbb{E}[r_n(\lambda)]} > \delta\right) \\
\leq &\delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-2}\mathbb{E}[(\mathbf{f}^T(\mathbf{I}_n - \mathbf{A}_n(\lambda))\varepsilon)^2] \\
= &\delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-2}\sigma^2\mathrm{tr}[(\mathbf{I}_n - \mathbf{A}_n(\lambda))^T\mathbf{f}\mathbf{f}^T(\mathbf{I}_n - \mathbf{A}_n(\lambda))] \\
= &\delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-2}\sigma^2\|(\mathbf{I}_n - \mathbf{A}_n(\lambda))^T\mathbf{f}\|^2 \\
= &\delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-1}\sigma^2\frac{\|(\mathbf{I}_n - \mathbf{A}_n(\lambda))^T\mathbf{f}\|^2}{n\mathbb{E}[r_n(\lambda)]} \\
\leq &\delta^2\sigma^2(n\mathbb{E}[r_n(\lambda)])^{-1} \to 0.
\end{aligned}$$

For (B.17.3), again using Chebyshev inequality, for any given $\delta > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{n^{-1}\left|\varepsilon^T\mathbf{A}_n(\lambda)\varepsilon - \sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)]\right|}{\mathbb{E}[r_n(\lambda)]} > \delta\right) \\
&\leq \delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-2}\mathbb{E}[(\varepsilon^T\mathbf{A}_n(\lambda)\varepsilon - \sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)])^2] \\
&= \delta^{-2}(n\mathbb{E}[r_n(\lambda)])^{-1}\frac{\mathbb{E}[(\varepsilon^T\mathbf{A}_n(\lambda)\varepsilon)^2] - (\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)])^2}{n\mathbb{E}[r_n(\lambda)]}.
\end{aligned}$$

Since $n\mathbb{E}[r_n(\lambda)] \geq \sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2]$, we only need to show the following

$$\frac{\mathbb{E}[(\varepsilon^T\mathbf{A}_n(\lambda)\varepsilon)^2] - (\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)])^2}{\sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)^2]} < \text{Constant}. \tag{B.17.5}$$

Denote $\mathbf{A}_n(\lambda) = (A_{ij})_{n\times n}$, then we have

$$\begin{aligned}
\mathbb{E}[(\varepsilon^T\mathbf{A}_n(\lambda)\varepsilon)^2] &= \mathbb{E}[(\sum_{i,j}A_{ij}\varepsilon_i\varepsilon_j)(\sum_{k,l}A_{kl}\varepsilon_k\varepsilon_l)] \\
&= \mathbb{E}[(\sum_{i}A_{ii}\varepsilon_i^2)(\sum_{k}A_{kk}\varepsilon_k^2)] + \mathbb{E}[(\sum_{i\neq j}A_{ij}\varepsilon_i\varepsilon_j)(\sum_{k\neq l}A_{kl}\varepsilon_k\varepsilon_l)] \\
&\leq \left(\sum_{i=1}^{n}A_{ii}\sigma^2\right)^2 + \sum_{i=1}^{n}A_{ii}^2\mathbb{E}[\varepsilon_i^4] + \sum_{i\neq j}A_{ij}^2\sigma^4.
\end{aligned}$$

There exists a constant c such that $\mathbb{E}[\varepsilon_i^4] \leq c\sigma^2$ and $\sigma^4 \leq c\sigma^2$, we have

$$\begin{aligned}\mathbb{E}[(\varepsilon^T \mathbf{A}_n(\lambda)\varepsilon)^2] &\leq \left(\sum_{i=1}^n A_{ii}\sigma^2\right)^2 + c\sum_{ij} A_{ij}\sigma^2 \\ &= (\sigma^2 \mathrm{tr}[\mathbf{A}_n(\lambda)])^2 + c\sigma^2 \mathrm{tr}[\mathbf{A}_n(\lambda)^2],\end{aligned}$$

which finishes our proof of (B.17.3).

For (B.17.4), using the proof of (B.17.2), (B.17.3) and $\sigma^2(n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda)])^2 \leq \sigma^2 n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda)^2] \leq \mathbb{E}[r_n(\lambda)]$, we only need to show

$$\sup_{\lambda>0} \frac{|\sigma^2 - n^{-1}\|\varepsilon\|^2|}{(\mathbb{E}[r_n(\lambda)])^{1/2}} \to 0, \tag{B.17.6}$$

since the fact that

$$\begin{aligned}&\left|\sigma^2 n^{-1}\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\lambda)] - n^{1-}\|(\mathbf{I}_n - \mathbf{A}_n(\lambda))\|^2\right| \\ &= \left|\sigma^2 - \sigma^2 n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda)] - n^{-1}\|\varepsilon + \mathbf{f} - \hat{\mathbf{f}}_n(\lambda)\|^2\right| \\ &= \left|\sigma^2 - \sigma^2 n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda)] - n^{-1}\|\varepsilon\|^2 - r_n(\lambda) - 2n^{-1}(\mathbf{f} - \hat{\mathbf{f}}_n(\lambda))^T\varepsilon\right| \\ &\leq \left|\sigma^2 - n^{-1}\|\varepsilon\|^2\right| + r_n(\lambda) + 2n^{-1}\left|\mathbf{f}^T(\mathbf{I}_n - \mathbf{A}_n(\lambda)\varepsilon\right| \\ &+ 2n^{-1}\left|\varepsilon^T\mathbf{A}_n(\lambda)\varepsilon - \sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)]\right| + \sigma^2 n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda)].\end{aligned}$$

By the Chebyshev's inequality, for any given $\delta > 0$, we have

$$\begin{aligned}&\mathbb{P}\left(\frac{|\sigma^2 - n^{-1}\|\varepsilon\|^2|}{(\mathbb{E}[r_n(\lambda)])^{1/2}} > \delta\right) \\ &\leq \delta^2(\mathbb{E}[r_n(\lambda)])^{-1}\mathbb{E}[(\sigma^2 - n^{-1}\|\varepsilon\|^2)^2] \\ &= \delta^2(\mathbb{E}[r_n(\lambda)])^{-1}\left(n^{-2}\mathbb{E}[\|\varepsilon\|^4] - \sigma^4\right) \\ &\leq \delta^2(\mathbb{E}[r_n(\lambda)])^{-1}\left(n^{-2}(n^2\sigma^4 + n\mathbb{E}[\varepsilon_i^4]) - \sigma^4\right) \\ &= \delta^2(n\mathbb{E}[r_n(\lambda)])^{-1}\mathbb{E}[\varepsilon_i^4] \to 0.\end{aligned}$$

Now it remains to prove (2.3.17), the numerator of which can be rearranged as

$$
\begin{aligned}
& n^{-1}\|\tilde{\mathbf{f}}_n(\hat{\lambda})\|^2 \\
&= \left(\frac{\sigma^2 n^{-1}\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\lambda)]}{n^{-1}\|(\mathbf{I}_n - \mathbf{A}_n(\lambda))\mathbf{y}\|^2} - 1\right)^2 n^{-1}\|(\mathbf{I}_n - \mathbf{A}_n(\lambda))\mathbf{y}\|^2 \\
&= \frac{((\sigma^2 - n^{-1}\|\varepsilon\|^2) - r_n(\lambda) - 2A_1 + 2A_2 + \sigma^2 n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda)])^2}{n^{-1}\|(\mathbf{I}_n - \mathbf{A}_n(\lambda))\mathbf{y}\|^2},
\end{aligned}
$$

where $A_1 = \left(n^{-1}\mathbf{f}^T(\mathbf{I}_n - \mathbf{A}_n(\lambda))\varepsilon\right)^2$ and $A_2 = n^{-1}(\varepsilon^T\mathbf{A}_n(\lambda)\varepsilon - \sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)])$.

It is not hard to see that

$$\frac{(\sigma^2 - n^{-1}\|\varepsilon\|^2)^2}{r_n(\lambda)} \to 0, \tag{B.17.7}$$

$$\frac{\left(n^{-1}\mathbf{f}^T(\mathbf{I}_n - \mathbf{A}_n(\lambda))\varepsilon\right)^2}{r_n(\lambda)} \to 0, \tag{B.17.8}$$

$$\frac{\left(n^{-1}(\varepsilon^T\mathbf{A}_n(\lambda)\varepsilon - \sigma^2\mathrm{tr}[\mathbf{A}_n(\lambda)])\right)^2}{r_n(\lambda)} \to 0, \tag{B.17.9}$$

$$\frac{(n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda)])^2}{r_n(\lambda)} \to 0. \tag{B.17.10}$$

based on the proofs of Lemma 2.3.13. □

B.18 Proof of Theorem 2.3.6

In order to prove Theorem 2.3.6, we need the following lemmas.

Lemma B.18.1. *Under the condition (A.2), we have $\tilde{r}_n(\lambda) \to 0$ when λ_n is from (A.2)*

Proof. To get $\tilde{r}_n(\lambda_n) \to 0$, it suffices to show that

$$\frac{\|(\mathbf{I}_n - \mathbf{A}_n(\lambda_n))\mathbf{y}\|^2}{\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\lambda_n)]} \to \sigma^2, \tag{B.18.1}$$

since the following derivation

$$\begin{aligned}
\tilde{r}_n(\lambda_n) &= n^{-1}\|\tilde{\mathbf{f}}_n(\lambda_n) - \mathbf{f}\|^2 \\
&= n^{-1}\|\varepsilon - \sigma^2\frac{\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\lambda_n)]}{\|(\mathbf{I}_n - \mathbf{A}_n(\lambda_n))\mathbf{y}\|^2}(\mathbf{I}_n - \mathbf{A}_n(\lambda_n))\mathbf{y}\|^2 \\
&= n^{-1}\|\varepsilon - \sigma^2\frac{\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\lambda_n)]}{\|(\mathbf{I}_n - \mathbf{A}_n(\lambda_n))\mathbf{y}\|^2}(\varepsilon + \mathbf{f} - \hat{\mathbf{f}}_n(\lambda_n))\mathbf{y}\|^2 \\
&\leq n^{-1}\left(1 - \sigma^2\frac{\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\lambda_n)]}{\|(\mathbf{I}_n - \mathbf{A}_n(\lambda_n))\mathbf{y}\|^2}\right)^2\|\varepsilon\|^2 \\
&+ 2n^{-1}\left|1 - \sigma^2\frac{\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\lambda_n)]}{\|(\mathbf{I}_n - \mathbf{A}_n(\lambda_n))\mathbf{y}\|^2}\right|\|\varepsilon\|\|\mathbf{f} - \hat{\mathbf{f}}_n(\lambda_n)\| \\
&+ n^{-1}\left(\sigma^2\frac{\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\lambda_n)]}{\|(\mathbf{I}_n - \mathbf{A}_n(\lambda_n))\mathbf{y}\|^2}\right)^2\|\mathbf{f} - \hat{\mathbf{f}}_n(\lambda_n)\|^2.
\end{aligned}$$

Obviously, (B.18.1) follows from Lemma 2.3.12. □

Lemma B.18.2. *Under the condition (A.2), we have* $\tilde{r}_n(\hat{\lambda}_G) \to 0$.

Proof. From the uniform consistency of $\mathrm{SURE}_n(\lambda)$ together with the fact that $\hat{\lambda}_G$ minimizes $\mathrm{SURE}_n(\lambda)$, we have

$$\begin{aligned}
\tilde{r}_n(\hat{\lambda}_G) &= \mathrm{SURE}_n(\hat{\lambda}_G) + o_p(1) \\
&\leq \mathrm{SURE}_n(\lambda_n) + o_p(1) \\
&= \tilde{r}_n(\lambda_n) + o_p(1) = o_p(1).
\end{aligned}$$

This is equivalent to say that $\tilde{r}_n(\hat{\lambda}_G) \to 0$. □

Lemma B.18.3. *Under the condition (A.2), we have* $\mathrm{GCV}_n(\hat{\lambda}_G) \to \sigma^2$.

Proof. This is trivial from Lemma B.18.2. □

Lemma B.18.4. *If* ε_i*'s are i.i.d* $N(0, \sigma^2)$*, for any* $\delta > 0$*, we have*

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{\|(\mathbf{I}_n - \mathbf{A}_n(\hat{\lambda}_G))\mathbf{y}\|^2}{\|(\mathbf{I}_n - \mathbf{A}_n(\hat{\lambda}_G))\mathbf{f}\|^2 + \sigma^2\mathrm{tr}[(\mathbf{I}_n - \mathbf{A}_n(\hat{\lambda}_G))^2]} \leq 1 - \delta\right) = 0. \qquad \text{(B.18.2)}$$

Proof. According to the proof of Lemma 5.2 in [58], the above lemma can be established directly. □

Lemma B.18.5. *For any sequence* $\{\lambda_n\}$ *such that* $\mathrm{GCV}_n(\lambda_n) \to \sigma^2$ *under the condition (A.3), we have* $n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda_n)] \to 0$.

Proof. Using Lemma B.18.4 and $\{\lambda_n\}$ such that $\mathrm{GCV}_n(\lambda_n) \to \sigma^2$, we have

$$\left(n^{-1}\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\hat{\lambda}_n)]\right)^2 \geq \left(n^{-1}\mathrm{tr}[(\mathbf{I}_n - \mathbf{A}_n(\hat{\lambda}_n))^2]\right)(1 - o_p(1)). \tag{B.18.3}$$

With the fact that $\left(n^{-1}\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\hat{\lambda}_n)]\right)^2 \leq n^{-1}\mathrm{tr}[(\mathbf{I}_n - \mathbf{A}_n(\hat{\lambda}_n))^2]$, we get

$$\frac{\left(n^{-1}\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\hat{\lambda}_n)]\right)^2}{n^{-1}\mathrm{tr}[(\mathbf{I}_n - \mathbf{A}_n(\hat{\lambda}_n))^2]} \to 1. \tag{B.18.4}$$

Recall that $\mathbf{A}_n(\hat{\lambda}) = (\mathbf{I}_n + \hat{\lambda}_n\mathbf{M})^{-1} = (\mathbf{I}_n + \mathbf{K}_n(\hat{\lambda}_n))^{-1}$ and $0 \leq \kappa_1 \leq \cdots \leq \kappa_n$ are the eigenvalues of $\mathbf{K}_n(\hat{\lambda}_n)$. It is clear that $\mathbf{I}_n - \mathbf{A}_n(\hat{\lambda}_n)$ have eigenvalues $\{\frac{\kappa_i}{1+\kappa_i}\}$. Similarly as in [58], let κ be the random variable taking values κ_i with probability n^{-1} for each $i \in \{1, 2, \cdots, n\}$. Then (B.18.4) means

$$\frac{\kappa(1+\kappa)^{-1}}{\mathbb{E}[\kappa(1+\kappa)^{-1}]} \to 1.$$

This implies that both $\kappa_{[pn]}(1+\kappa_{[pn]})^{-1}$ and $\kappa_{[qn]}(1+\kappa_{[qn]})^{-1}$ tend to $\mathbb{E}[\kappa(1+\kappa)^{-1}]$, we have $\mathbb{E}[\kappa(1+\kappa)^{-1}] \to 1$, from which $n^{-1}\mathrm{tr}[\mathbf{A}_n(\hat{\lambda}_n)] \to 0$ follows. □

Lemma B.18.6. *For sequence* $\{\lambda_n\}$ *such that* $\mathrm{GCV}_n(\lambda_n) \to \sigma^2$, $\hat{\mathbf{f}}_n(\lambda_n)$ *is consistent if and only if* $n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda_n)] \to 0$.

Proof. If $\hat{\mathbf{f}}_n(\lambda_n)$ is consistent, $r_n(\lambda_n) \to 0$ and hence $n^{-1}\|\mathbf{y} - \hat{f}_n(\lambda_n)\|^2 \to \sigma^2$ since $n^{-1}\|\varepsilon\|^2 \to \sigma^2$. Then from the fact that $\mathrm{GCV}_n(\lambda_n) = \frac{n^{-1}\|\mathbf{y}-\hat{\mathbf{f}}_n(\lambda_n)\|^2}{(n^{-1}\mathrm{tr}[\mathbf{I}_n-\mathbf{A}_n(\lambda_n)])^2} \to \sigma^2$, we have $\left(n^{-1}\mathrm{tr}[\mathbf{I}_n - \mathbf{A}_n(\lambda_n)]\right)^2 \to 1$ and thus $n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda_n)] \to 0$.

Conversely, if $n^{-1}\mathrm{tr}[\mathbf{A}_n(\lambda_n)] \to 0$, since $\mathrm{GCV}_n(\lambda_n) \to \sigma^2$, we have $n^{-1}\|\mathbf{y} - \hat{f}_n(\lambda_n)\|^2 \to \sigma^2$. Then with the fact that $n^{-1}\|\varepsilon\|^2 \to \sigma^2$, we have $r_n(\lambda_n) \to 0$, which implies that $\hat{\mathbf{f}}_n(\lambda_n)$ is consistent. □

From Lemmas B.18.3, B.18.5 and B.18.6, Theorem 2.3.6 is proved.

B.19 Proof of Theorem 2.3.7

Proof. From the condition (A.2), for λ_n^* which is the minimizer of $r_n(\lambda)$, we have $r_n(\lambda_n^*) \to 0$. According to Lemma 2.3.13 we have

$$\frac{(n^{-1}\mathrm{tr}[\mathbf{A_n}(\lambda_\mathbf{n}^*)])^\mathbf{2}}{n^{-1}\mathrm{tr}[\mathbf{A_n}(\lambda_\mathbf{n}^*)^\mathbf{2}]} \to 0.$$

Hence from Lemma 2.3.14, we know $\mathrm{SURE}_n(\lambda_n^*) - n^{-1}\|\varepsilon_n\|^2 + \sigma^2 = r_n(\lambda_n^*)(1 + o_p(1))$.

On the other hand, from Theorem 2.3.6 this also holds for $\hat{\lambda} = \hat{\lambda_G}$. Therefore we have

$$\mathrm{SURE}_n(\hat{\lambda}_G) - n^{-1}\|\varepsilon_n\|^2 + \sigma^2 = r_n(\hat{\lambda}_G)(1 + o_p(1)).$$

Since $\mathrm{SURE}_n(\hat{\lambda}_G) \leq \mathrm{SURE}_n(\lambda_n^*)$ and $r_n(\lambda_n^*) \leq r_n(\hat{\lambda}_G)$, we have $r_n(\hat{\lambda}_G)/r_n(\lambda_n^*) \xrightarrow{p} 1$. □

B.20 Proof of Theorem 2.3.4

Proof. Without loss of generality, let $\mathbf{f}$ be the vector of function values at the knots of $T = \{X_i\}_{i=1}^n$ normalized by $|f|_{T,0}^2 = \frac{1}{n}\mathbf{f^T f} = \mathbf{1}$.

As in the proof of 2.3.3,

$$\begin{aligned}
&\left|\frac{1}{nVt^{d/2+1}}\frac{\mathbf{f^T L f}}{\mathbf{f^T f}}-\frac{|f|^2_{T,1}}{|f|^2_{T,0}}\right| \\
=&\left|\frac{1}{n^2Vt^{d/2+1}}\mathbf{f^T L f}-|\mathbf{f}|^{\mathbf{2}}_{\mathbf{T,1}}\right| \\
=&\left|\frac{1}{n^2Vt^{d/2+1}}\sum_{i=1}^{n}f(x_i)\sum_{j:j\neq i}(w_{i,j}f(x_i)-w_{i,j}f(x_j))-\frac{1}{n}\sum_{i=1}^{n}f(x_i)\Delta f(x_i)\right| \\
\leq&\frac{M}{n}\sum_{i=1}^{n}\left|\frac{1}{nVt^{d/2+1}}\sum_{j:j\neq i}(w_{i,j}f(x_i)-w_{i,j}f(x_j))-\Delta f(x_i)\right|,
\end{aligned}$$

where M is the upper bound of $|f(X)|$ on Ω.

According to Theorem 3.1 in [56], we have $|\mu_j - e_j| \to 0$ as $n \to \infty$. Hence from Theorem 2.3.2, we have

$$C_9 j^{2/d} \leq \mu_j \leq C_{10} j^{2/d}.$$

For the case $m > 1$, we have the following clearly

$$C_9 j^{2m/d} \leq \mu_j \leq C_{10} j^{2m/d},$$

which completes the proof. □

B.21 Agmon's Theorem

Let $A(x, D) = \sum_{|\alpha|\leq m'} a_\alpha(x) D^\alpha$ be an elliptic operator of order m' in Ω, having continuous leading coefficients and bounded, measurable lower order coefficients. Let A be symmetric over $C_0^\infty(\Omega)$ in the sense that for all $\phi, \psi \in C_0^\infty(\Omega), (A\phi, \psi)_{0,\Omega} = (\phi, A\psi)_{0,\Omega}$. Suppose that there exists an unbounded self-adjoint transformation $\mathcal{A}$ on $L_2(\Omega)$, such that $C_0^\infty(\Omega) \subset D(\mathcal{A}) \subset H_{m'}(\Omega)$ and $\mathcal{A}u = Au, \forall u \in D(\mathcal{A})$. Let $n' = n$ if n is odd, $n' = n+1$ if n is even. In case $m' \leq n'$, suppose that there exists an odd positive integer k such that $k > n'/m'$, the coefficients of A are in $C^{(k-1)m'*}(\Omega)$ and $D(\mathcal{A}^k) \subset H_{km'}(\Omega)$.

Then the spectrum of $\mathcal{A}$ is discrete, and the eigenvalues of $\mathcal{A}$ have finite multiplicity. Let $\{\lambda_j\}$ be the sequence of eigenvalues of $\mathcal{A}$ counted according to multiplicity. For $\lambda > 0$, let $N_+(\lambda)$ be the number of nonnegative eigenvalues $\lambda_j \leq \lambda$, then

$$N_+(\lambda) = c\lambda^{n/m'} + o(\lambda^{n/m'}),$$

as $\lambda \to \infty$, where $c = (2\pi)^{-n} \int_\Omega w(x)dx$, and $w(x) = |\{\xi : 0 < A'(x, i\xi) < 1\}|$. Note that $A'(x, \xi) = \sum_{|\alpha|=m'} a_\alpha(x)\xi^\alpha$.

B.22 Neumann Boundary Condition

Consider the following physical problem: a planar object is surrounded by material capable of transferring heat at a prescribed rate $f(x, y)$; our objective is to find the equilibrium temperature inside the object. The corresponding PDE problem is as follow: let Ω be a closed region of the plane, find the function $\phi(x, y)$ such that

$$\begin{aligned} \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} &= 0 \quad \forall (x, y) \in \Omega, \\ \frac{\partial\phi}{\partial \mathbf{n}} &= kf(x, y) \quad \forall (x, y) \in \partial\Omega, \end{aligned}$$

where $\mathbf{n}$ is the normal direction to the boundary. Such a PDE bondary value problem is called *Neumann problem*. It is also obvious that the physical problem is ill-posed unless the total rate at which heat flows into the object is 0, which requires the condition

$$\int_{\partial\Omega} \frac{\partial\phi}{\partial \mathbf{n}} ds = \int_{\partial\Omega} f(x, y)ds = 0.$$

We consider a sequence of queueing systems indexed by n. It is assumed that each system is composed of J stations, indexed by 1 through J, and K customer classes, indexed by 1 through K. Each customer class has a fixed route through the network of stations. Customers in class k, $k = 1, \ldots, K$, arrive to the system according to a renewal process,

independently of the arrivals of the other customer classes. These customers move through the network, never visiting a station more than once, until they eventually exit the system.

APPENDIX C

PROOFS IN CHAPTER 3

C.1 Proofs in Section 3.3.2

Throughout this proof, we will assume that the variance parameter in Formulation (3.3.1): $\sigma^2 = 1$. We will also assume that the observed feature matrix X is standardized with column mean as 0 and standard deviation in the order of $\sqrt{n}$. The general case would follow from a simple rescaling procedure. Let $\mathcal{S}_0 := supp(\beta^*)$, the support set of the ground truth β^*.

C.1.1 Proof of Lemma 3.3.1

Proof. Given the data generation mechanism in (3.3.1), the random error vector ϵ is independent sub-Gaussian random variables. We thus have the following tail inequality:

$$\mathbb{P}(\frac{\epsilon_i}{\sqrt{n}} > t) \leq 2e^{-\frac{nt^2}{2}}.$$

By Bonferroni bound, we have

$$\mathbb{P}(\|\frac{\epsilon_i}{\sqrt{n}}\|_\infty > t) \leq 2ne^{-\frac{nt^2}{2}}.$$

Let $t = O(\sqrt{\frac{\log p}{n}})$, say $\lambda = t = \sqrt{\alpha\frac{\log p}{n}}$ for some $\alpha > 2$, we have with probability larger than $1 - \frac{2n}{p} \cdot (\frac{1}{p})^{\frac{\alpha-2}{2}}$,

$$|\frac{\epsilon_i}{\sqrt{n}}| = \frac{1}{\sqrt{n}}|Y_i - \langle (X^T)_i, \beta^* \rangle| \leq \lambda,$$

for all $i = 1, \cdots, n$.

□

C.1.2 Proof of Theorem 3.3.1

Proof. In order to prove the statement, we will need the following lemmas.

Lemma C.1.1. *Let $\nu = \hat{\beta} - \beta^*$, the difference vector of the optimal solution and the ground truth. We have $\nu \in \{x \in \mathbb{R}^p : \|x_{\mathcal{S}_0^c}\|_1 \leq \|x_{\mathcal{S}_0}\|_1\}$, which is a cone in $\mathbb{R}^p$.*

Proof. Due to the definition of $\hat{\beta}$, which is the optimal solution to (3.3.4), and β^* is a feasible solution to (3.3.4), we have

$$\|\hat{\beta}\|_1 \leq \|\beta^*\|_1.$$

Thus, we have

$$\begin{aligned}
\|\beta^*\|_1 &\geq \|\hat{\beta}\|_1 \\
&= \|\beta^* + \nu\|_1 \\
&= \|(\beta^* + \nu)_{\mathcal{S}_0}\|_1 + \|(\beta^* + \nu)_{\mathcal{S}_0^c}\|_1 \\
&\geq \|\beta^*_{\mathcal{S}_0}\|_1 - \|\nu_{\mathcal{S}_0}\|_1 + \|\nu_{\mathcal{S}_0^c}\|_1 \\
&\geq \|\beta^*\|_1 - \|\nu_{\mathcal{S}_0}\|_1 + \|\nu_{\mathcal{S}_0^c}\|_1,
\end{aligned}$$

which proves the lemma. □

According to the restricted eigenvalue assumption, we have the following inequality:

$$\gamma\|\nu\|_2^2 \leq \frac{1}{n}\|X\nu\|_2^2. \tag{C.1.1}$$

On the other hand,

$$\begin{aligned}
\frac{1}{n}\|X\nu\|_2^2 &= \frac{1}{n}\nu^T X^T X\nu \\
&\leq \frac{1}{n}\|X^T X\nu\|_\infty \|\nu\|_1.
\end{aligned} \tag{C.1.2}$$

Since $\|\nu_{\mathcal{S}_0^c}\|_1 \leq \|\nu_{\mathcal{S}_0}\|_1$, and according to Cauchy–Schwarz inequality, we have

$$\|\nu_{\mathcal{S}_0^c}\|_1 \leq \|\nu_{\mathcal{S}_0}\|_1 \leq \sqrt{S}\|\nu_{\mathcal{S}_0}\|_2.$$

Thus

$$\begin{aligned} \|\nu\|_1 &= \|\nu_{\mathcal{S}_0^c}\|_1 + \|\nu_{\mathcal{S}_0}\|_1 \\ &\leq 2\|\nu_{\mathcal{S}_0}\|_1 \\ &\leq 2\sqrt{S}\|\nu_{\mathcal{S}_0}\|_2 \\ &\leq 2\sqrt{S}\|\nu\|_2. \end{aligned} \tag{C.1.3}$$

For $\frac{1}{n}\|X^T X\nu\|_\infty$, we have

$$\begin{aligned} \frac{1}{n}\|X^T X\nu\|_\infty &= \max_{i=1}^{p} |\langle \frac{1}{\sqrt{n}} X_i, \frac{1}{\sqrt{n}} X\nu \rangle| \\ &\leq \max_{i=1}^{p} \|\frac{1}{\sqrt{n}} X_i\|_1 \|\frac{1}{\sqrt{n}} X\nu\|_\infty \\ &\leq \|\frac{1}{\sqrt{n}} X\|_1 \|\frac{1}{\sqrt{n}} X\nu\|_\infty. \end{aligned} \tag{C.1.4}$$

According to our Formulation (3.3.4), and the fact that both $\hat{\beta}$ and β^* are feasible to (3.3.4), we have

$$\|\frac{1}{\sqrt{n}} X\nu\|_\infty \leq 2\lambda.$$

According to the normalization on X and Cauchy–Schwarz inequality, we have

$$\|\frac{1}{\sqrt{n}} X\|_1 \leq O(\sqrt{n}). \tag{C.1.5}$$

Combing the upper-bounds above, we have

$$\begin{aligned} \gamma\|\nu\|_2^2 \leq \frac{1}{n}\|X\nu\|_2^2 &\leq \frac{1}{n}\|X^T X\nu\|_\infty \|\nu\|_1 \\ &\leq O(\lambda\sqrt{Sn})\|\nu\|_2. \end{aligned} \tag{C.1.6}$$

Thus we obtain that

$$\|\nu\|_2 \leq O(\sqrt{S \log p}). \tag{C.1.7}$$

□

C.1.3 Proof of Theorem 3.3.2

Proof. Recall the way we bound $\|\frac{1}{\sqrt{n}}X\|_1$ in Inequality (C.1.5), if we in addition have that $\|X\|_1 = O(\sqrt{n})$, we have

$$\begin{aligned} \gamma\|\nu\|_2^2 \leq \frac{1}{n}\|X\nu\|_2^2 &\leq \frac{1}{n}\|X^T X\nu\|_\infty \|\nu\|_1 \\ &\leq O(\lambda\sqrt{S})\|\nu\|_2. \end{aligned} \tag{C.1.8}$$

Thus we obtain that

$$\|\nu\|_2 \leq O(\sqrt{\frac{S \log p}{n}}). \tag{C.1.9}$$

□

C.1.4 Proof of Theorem 3.3.3

Proof. The proof simply follows from the previous proofs:

$$\begin{aligned} \frac{1}{n}\|X\nu\|_2^2 &\leq \frac{1}{n}\|X^T X\nu\|_\infty \|\nu\|_1 \\ &\leq O(\lambda\sqrt{S})\|\nu\|_2 \\ &= O(\frac{S \log p}{n}). \end{aligned} \tag{C.1.10}$$

□

C.2 One useful proposition

In our simulation, we used the following results to

Proposition C.2.1. *If we choose the tuning parameter λ^* in the Lagrangian Formulation (3.3.6) such that $\lambda > \|X\|_\infty$, there exists one unique global minimum to (3.3.6), with $\hat{\beta} = 0$.*

C.2.1 Proof of Proposition C.2.1

Proof. We will start with the simplest case where $p = 1$. In this case, we have the objective as

$$\min \|Y - X\beta\|_\infty + \lambda|\beta|, \tag{C.2.1}$$

where $Y, X \in \mathbb{R}^n$, $\beta \in \mathbb{R}$. By the definition of the ℓ_∞ norm, we can further write the objective in (C.2.1) as

$$\begin{aligned} &\min \|Y - X\beta\|_\infty + \lambda|\beta| \\ =&\min \max_{i=1,\cdots,n} \{|Y_i - X_i\beta|\} + \lambda|\beta|. \end{aligned}$$

It can be directly seen that this is a convex problem. For the first part $\max_{i=1,\cdots,n}\{|Y_i - X_i\beta|\}$, which is the maximum of a set of convex functions, the resulting function is still convex. The second part $\lambda|\beta|$ itself is convex. Finding the minimum to (C.2.1) is equivalent to find a stationary solution to it.

On the one hand, for any $\beta_0 > 0$, we have $\lambda|\beta| = \lambda\beta$. Let $i \in I = \{i : |Y_i - X_i\beta_0| = \|Y - X\beta_0\|_\infty\}$, we have in a small neighborhood of β_0, $\|Y - X\beta\|_\infty = |Y_i - X_i\beta|$. Thus in this small neighborhood $[\beta_-, \beta_+]$, where $\beta_-, \beta_+ \geq 0$, the objective is simply

$$f(\beta) = |Y_i - X_i\beta| + \lambda\beta.$$

By first order condition of the above function $f(\beta)$, we have

$$\partial f(\beta) = \lambda - sign(Y_i - X_i\beta)X_i.$$

According to the assumption on λ, the function $f(\beta)$ is increasing in the small neighborhood $[\beta_-, \beta_+]$.

On the other hand, for any $\beta_0 < 0$, we have $\lambda|\beta| = -\lambda\beta$. Let $i \in I = \{i : |Y_i - X_i\beta_0| = \|Y - X\beta_0\|_\infty\}$, we have in a small neighborhood of β_0, $\|Y - X\beta\|_\infty = |Y_i - X_i\beta|$. Thus in this small neighborhood $[\beta_-, \beta_+]$, where $\beta_-, \beta_+ \leq 0$, the objective is simply

$$f(\beta) = |Y_i - X_i\beta| - \lambda\beta.$$

By first order condition of the above function $f(\beta)$, we have

$$\partial f(\beta) = -\lambda - sign(Y_i - X_i\beta)X_i.$$

According to the assumption on λ, the function $f(\beta)$ is decreasing in the small neighborhood $[\beta_-, \beta_+]$.

Using the two observations above, we can obtain that $\beta = 0$ is the unique global minimum for the objective in (C.2.1).

Now, we will generalize the above proof in the 1D case to high-dimensional case, where we assume $\beta \in \mathbb{R}^p$, $p > 1$.

For any $\beta_0 \in \mathbb{R}^p$, let $J = \{j : \beta_{0j} \neq 0\}$. Define the neighborhood of β_0 as $\mathcal{N} = \{\beta \in \mathbb{R}^p : \beta_j = 0, \text{ for } j \notin J; \beta_j \in [\beta_{0j} - \epsilon, \beta_{0j} + \epsilon], \text{ for } j \in J\}$, where $\epsilon > 0$ is chosen such that $(\beta_{0j} - \epsilon)(\beta_{0j} + \epsilon) > 0$ for all $j \in J$.

Let $i \in I = \{i : |Y_i - X_i\beta_0| = \|Y - X\beta_0\|_\infty\}$, we have that in the small neighborhood $\mathcal{N}$ of β_0, $\|Y - X\beta\|_\infty = |Y_i - X_i\beta|$. Thus within this small neighborhood $\mathcal{N}$, the objective

is simply

$$f(\beta) = |Y_i - X_i\beta| + \lambda\|\beta\|_1.$$

By first order condition of the above function $f(\beta)$, we have

$$\partial f(\beta) = \lambda sign(\beta) - sign(Y_i - X_i\beta)X_i,$$

where $sign(\beta) \in \mathbb{R}^p$ is the indicator vector of the signs for β. According to the assumption on λ, $(\partial f(\beta))_k > 0$ for $k \in \mathbb{K}_1 = \{k : \beta_{0k} > 0\}$; $(\partial f(\beta))_k < 0$ for $k \in \mathbb{J}/\mathbb{K}_1$. Thus we conclude that $\beta = 0$ is the unique global minimum for the objective in (3.3.4). □

REFERENCES

[1] Miju Ahn, Jong-Shi Pang, and J Xin. "Difference-of-convex learning I: Directional stationarity, optimality, and sparsity". In: *SIAM Journal on Optimization, revision under review (as of February 2017)* (2016).

[2] C. J. Stone. "Optimal global rates of convergence for nonparamteric regression". In: *The Annals of Statistics* 10 (1982), pp. 1040–1053.

[3] Peng Zhao and Bin Yu. "On model selection consistency of Lasso". In: *Journal of Machine learning research* 7.Nov (2006), pp. 2541–2563.

[4] Martin J Wainwright. "Sharp thresholds for High-Dimensional and noisy sparsity recovery using l_1-Constrained Quadratic Programming (Lasso)". In: *IEEE transactions on information theory* 55.5 (2009), pp. 2183–2202.

[5] Jianqing Fan and Runze Li. "Variable selection via nonconcave penalized likelihood and its oracle properties". In: *Journal of the American statistical Association* 96.456 (2001), pp. 1348–1360.

[6] Jianqing Fan and Heng Peng. "Nonconcave penalized likelihood with a diverging number of parameters". In: *The Annals of Statistics* 32.3 (2004), pp. 928–961.

[7] Hui Zou. "The adaptive lasso and its oracle properties". In: *Journal of the American statistical association* 101.476 (2006), pp. 1418–1429.

[8] Jian Huang, Shuangge Ma, and Cun-Hui Zhang. "Adaptive Lasso for sparse high-dimensional regression models". In: *Statistica Sinica* (2008), pp. 1603–1618.

[9] Cun-Hui Zhang. "Nearly unbiased variable selection under minimax concave penalty". In: *The Annals of statistics* 38.2 (2010), pp. 894–942.

[10] Tong Zhang. "Analysis of multi-stage convex relaxation for sparse regularization". In: *Journal of Machine Learning Research* 11.Mar (2010), pp. 1081–1107.

[11] Tong Zhang. "Multi-stage convex relaxation for feature selection". In: *Bernoulli* 19.5B (2013), pp. 2277–2293.

[12] Emmanuel Candes, Terence Tao, et al. "The Dantzig selector: Statistical estimation when p is much larger than n". In: *The annals of Statistics* 35.6 (2007), pp. 2313–2351.

[13] Peter J Bickel, Ya'acov Ritov, and Alexandre B Tsybakov. "Simultaneous analysis of Lasso and Dantzig selector". In: *The Annals of Statistics* (2009), pp. 1705–1732.

[14] Peter Bühlmann and Sara Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

[15] Robert Tibshirani. "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), pp. 267–288.

[16] Scott Chen and David L Donoho. "Examples of basis pursuit". In: *SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation*. International Society for Optics and Photonics. 1995, pp. 564–574.

[17] Cun-Hui Zhang and Stephanie S Zhang. "Confidence intervals for low dimensional parameters in high dimensional linear models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.1 (2014), pp. 217–242.

[18] Adel Javanmard and Andrea Montanari. "Hypothesis testing in high-dimensional regression under the gaussian random design model: Asymptotic theory". In: *IEEE Transactions on Information Theory* 60.10 (2014), pp. 6522–6554.

[19] Sara Van de Geer et al. "On asymptotically optimal confidence regions and tests for high-dimensional models". In: *The Annals of Statistics* 42.3 (2014), pp. 1166–1202.

[20] Adel Javanmard and Andrea Montanari. "Confidence intervals and hypothesis testing for high-dimensional regression." In: *Journal of Machine Learning Research* 15.1 (2014), pp. 2869–2909.

[21] Po-Ling Loh and Martin J Wainwright. "Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima". In: *Advances in Neural Information Processing Systems*. 2013, pp. 476–484.

[22] Zhaoran Wang, Han Liu, and Tong Zhang. "Optimal computational and statistical rates of convergence for sparse nonconvex learning problems". In: *Annals of statistics* 42.6 (2014), p. 2164.

[23] Shuai Zhang and Jack Xin. "Minimization of Transformed L_1 Penalty: Theory, Difference of Convex Function Algorithm, and Robust Application in Compressed Sensing". In: *arXiv preprint arXiv:1411.5735* (2014).

[24] Jinchi Lv and Yingying Fan. "A unified approach to model selection and sparse recovery using regularized least squares". In: *The Annals of Statistics* (2009), pp. 3498–3528.

[25] Rahul Mazumder, Jerome H Friedman, and Trevor Hastie. “Sparsenet: Coordinate descent with nonconvex penalties”. In: *Journal of the American Statistical Association* 106.495 (2011), pp. 1125–1138.

[26] Hui Zou and Runze Li. “One-step sparse estimates in nonconcave penalized likelihood models”. In: *Annals of statistics* 36.4 (2008), p. 1509.

[27] AD Aleksandrov. “Surfaces represented as a difference of two convex functions, Russian Acad. Sci”. In: *Dokl. Math*. 1. 1950.

[28] Philip Hartman. “On functions representable as a difference of convex functions”. In: *Pacific Journal of Mathematics* 9.3 (1959), pp. 707–713.

[29] Reiner Horst and Nguyen V Thoai. “DC programming: overview”. In: *Journal of Optimization Theory and Applications* 103.1 (1999), pp. 1–43.

[30] J-B Hiriart-Urruty. “Generalized Differentiability/Duality and Optimization for Problems Dealing with Differences of Convex Functions”. In: *Convexity and duality in optimization*. Springer, 1985, pp. 37–70.

[31] Pham Dinh Tao and Le Thi Hoai An. “Convex analysis approach to dc programming: Theory, algorithms and applications”. In: *Acta Mathematica Vietnamica* 22.1 (1997), pp. 289–355.

[32] Hoang Tuy. “Global minimization of a difference of two convex functions”. In: *Nonlinear Analysis and Optimization* (1987), pp. 150–182.

[33] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

[34] Maher Nouiehed, Jong-Shi Pang, and Meisam Razaviyayn. “On the Pervasiveness of Difference-Convexity in Optimization and Statistics”. In: *arXiv preprint arXiv:1704.03535* (2017).

[35] Lan Wang, Yongdai Kim, and Runze Li. “Calibrating non-convex penalized regression in ultra-high dimension”. In: *Annals of statistics* 41.5 (2013), p. 2505.

[36] EL Lehmann and G Casella. “Theory of Point Estimation, Springer-Verlag”. In: *New York* (1998).

[37] Bharath K Sriperumbudur and Gert RG Lanckriet. “A proof of convergence of the concave-convex procedure using zangwill’s theory”. In: *Neural computation* 24.6 (2012), pp. 1391–1407.

[38] Pham Dinh Tao et al. “The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems”. In: *Annals of operations research* 133.1-4 (2005), pp. 23–46.

[39] Alan L Yuille and Anand Rangarajan. “The concave-convex procedure”. In: *Neural computation* 15.4 (2003), pp. 915–936.

[40] Jong-Shi Pang, Meisam Razaviyayn, and Alberth Alvarado. “Computing B-stationary points of nonsmooth DC programs”. In: *Mathematics of Operations Research* 42.1 (2016), pp. 95–118.

[41] Bradley Efron et al. “Least angle regression”. In: *The Annals of statistics* 32.2 (2004), pp. 407–499.

[42] Jianqing Fan, Lingzhou Xue, and Hui Zou. “Strong oracle optimality of folded concave penalized estimation”. In: *Annals of statistics* 42.3 (2014), p. 819.

[43] Alexander J Smola and Risi Kondor. “Kernels and regularization on graphs”. In: *Learning theory and kernel machines*. Springer, 2003, pp. 144–158.

[44] Mikhail Belkin, Irina Matveeva, and Partha Niyogi. “Regularization and semi-supervised learning on large graphs”. In: *International Conference on Computational Learning Theory*. Springer. 2004, pp. 624–638.

[45] Rie Johnson and Tong Zhang. “On the effectiveness of Laplacian normalization for graph semi-supervised learning”. In: *Journal of Machine Learning Research* 8.Jul (2007), pp. 1489–1517.

[46] Rie K Ando and Tong Zhang. “Learning on graph with Laplacian regularization”. In: *Advances in neural information processing systems*. 2007, pp. 25–32.

[47] J. Huang et al. “The Sparse Laplacian Shrinkage Estimator for High-Dimensional Regression”. In: *The Annals of Statistics* 39.4 (2011), pp. 2021–2046.

[48] X. Zhou and M. Belkin. “Semi-supervised Learning by Higher Order Regularization”. In: *14th International Conference on Artificial Intelligence and Statistics* 15 (2011).

[49] Eric D Kolaczyk and Gábor Csárdi. *Statistical analysis of network data with R*. Vol. 65. Springer, 2014.

[50] Alisa Kirichenko, Harry van Zanten, et al. “Estimating a smooth function on a large graph by Bayesian Laplacian regularisation”. In: *Electronic Journal of Statistics* 11.1 (2017), pp. 891–915.

[51] Alisa Kirichenko, Harry van Zanten, et al. “Minimax lower bounds for function estimation on graphs”. In: *Electronic Journal of Statistics* 12.1 (2018), pp. 651–666.

[52] J. Duchon. “Splines minimizing rotation invariant seminorms in Sobolev spaces”. In: *Constructive Theory of Functions of Several Variables* Berlin:Springer-Verlag (1977), pp. 85–100.

[53] Simon N. Wood, Mark V. Bravington, and Sharon L. Hedley. “Soap film smoothing”. In: *Journal Of The Royal Statistical Society Series B* 70.5 (2008), pp. 931–955.

[54] G. Wahba. *Spline Models for Observational Data*. Vol. 59. CBMS-NSE Regional Conference Series in Applied Mathematics. Philadelphia: SIAM, 1990.

[55] T. Ramsay. “Spline smoothing over difficult regions”. In: *Journal of the Royal Statistical Society: Series B(Statistical Methodology)* 64 (2002), pp. 307–319.

[56] M. Belkin and P. Niyogi. “Towards a theorectical foundation for laplacian-based manifold methods”. In: *Journal of Computer and System Sciences* 74.8 (2008), pp. 1289–1308.

[57] Xueyuan Zhou and Nathan Srebro. “Error analysis of Laplacian eigenmaps for semi-supervised learning”. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. 2011, pp. 901–908.

[58] K. C. Li. “From Stein’s unbiased risk estimates to the method of generalized cross validation”. In: *The Annals of Statistics* 13 (1985), pp. 1352–1377.

[59] M. Belkin and P. Niyogi. “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation”. In: *Neural Computation* 15.6 (2003), pp. 1373–1396.

[60] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2001.

[61] Alexander Grigor’yan. “Heat kernels on weighted manifolds and applications”. In: *Cont. Math* 398.2006 (2006), pp. 93–191.

[62] S. Agmon. *Lectures on Elliptic Boundary Value Problems*. Princeton, 1965.

[63] Krishan K Chawla. *Composite materials: science and engineering*. Springer Science & Business Media, 2012.

[64] Dominic Gates. “Boeing finds 787 pieces arent quite a perfect fit”. In: *Rapport technique, Seattle Times* (2007).

[65] Yuchen Wen et al. “Feasibility analysis of composite fuselage shape control via finite element analysis”. In: *Journal of manufacturing systems* 46 (2018), pp. 272–281.

[66] Xiaowei Yue et al. “Surrogate Model-Based Control Considering Uncertainties for Composite Fuselage Assembly”. In: *Journal of Manufacturing Science and Engineering* 140.4 (2018), p. 041017.

[67] J. Du et al. “Optimal Placement of Actuators Using Sparse Learning for Composite Fuselage Shape Control”. In: *ASME Transactions, Journal of Manufacturing Science and Engineering* (under review).

[68] Haili Chui and Anand Rangarajan. “A new algorithm for non-rigid point matching”. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*. Vol. 2. IEEE. 2000, pp. 44–51.

[69] F James. “Fitting tracks in wire chambers using the Chebyshev norm instead of least squares”. In: *Nuclear Instruments and Methods in Physics Research* 211.1 (1983), pp. 145–152.

[70] Ali Zolghadri and David Henry. “Minimax statistical models for air pollution time series. Application to ozone time series data measured in Bordeaux”. In: *Environmental Monitoring and Assessment* 98.1-3 (2004), pp. 275–294.

[71] Chong Qi. “Theoretical uncertainties of the Duflo–Zuker shell-model mass formulae”. In: *Journal of Physics G: Nuclear and Particle Physics* 42.4 (2015), p. 045104.

[72] Mario Milanese and Gustavo Belforte. “Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: Linear families of models and estimators”. In: *IEEE Transactions on automatic control* 27.2 (1982), pp. 408–414.

[73] Alin Alecu et al. “Wavelet-based scalable L-infinity-oriented compression”. In: *IEEE Transactions on Image Processing* 15.9 (2006), pp. 2499–2512.

[74] E. Castillo, C. Castillo, and A.S. et al. Hadi. “Combined regression models”. In: *Computational Statistics* 24 (2009), pp. 37–66.

[75] KEITH KNIGHT. “On the asymptotic distribution of the ℓ_∞ estimator in linear regression”. In: 2017, manuscript.

[76] Gareth M James, Peter Radchenko, and Jinchi Lv. “DASSO: connections between the Dantzig selector and lasso”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.1 (2009), pp. 127–142.

[77] Emmanuel Candes and Terence Tao. “Decoding by linear programming”. In: *arXiv preprint math/0502327* (2005).

[78] James W Demmel. *Applied numerical linear algebra*. Vol. 56. Siam, 1997.

[79] B. A. Dye, X. Li, and E. D. Beltran-Aguilar. “Selected oral health indicators in the United States, 2005-2008”. In: *NCHS Data Brief* 96 (2012), pp. 1–8.

[80] US Department of Health and Human Services. *Oral health in America: a report of the Surgeon General*. Report. US DHHS, National Institute of Dental and Craniofacial Research, National Institutes of Health, 2000.

[81] Institute of Medicine. *Advancing oral health in America: The role of HHS*. Report. 2011.

[82] US Department of Health and Human Services. “Healthy people 2020, Topics and objectives, Oral health.” In: (2013).

[83] S.M. Skillman et al. “The challenge to delivering oral health services in rural America”. In: *Journal of public health dentistry* 70.s1 (2010), S49–S57.

[84] *Littrell A. Arkansas ConnectCare. In: Centers for Medicare & Medicaid Services, editor.* Government Document. 2012.

[85] *Roddy T, Tucker S. Maryland Healthy Smiles Dental Program. In: Centers for Medicare & Medicaid Services, editor.* Government Document. 2012.

[86] *State Medicaid and CHIP Program Support of Sustainable Oral Health Care Delivery Models in Schools and Community-Based Settings.* Government Document. 2014.

[87] Shanshan Cao et al. “Disparities in Preventive Dental Care Among Children in Georgia.” In: *Prev Chronic Dis* 14 (2017), pp. 170–176.

[88] “State Standards For Access To Care In Medicaid Managed Care”. In: *Department of Health and Human Services* (2014).

[89] *Georgia’s Rural Counties*. Government Document. 2014.

[90] S. Cao et al. “Identifying Shortage Areas for Preventive Dental Care for Children using High Geographic Granularity Estimates of Need and Supply”. In: *Public Health Reports* under 2nd review (2016).

[91] American Academy of Pediatric Dentistry. “Guideline on periodicity of examination, preventive dental services, anticipatory guidance/counseling, and oral treat-

ment for infants, children, and adolescents". In: *Pediatr Dent* 30.7 Suppl (2008), pp. 112–8.

[92] American Academy of Pediatric Dentistry. "2010 Survey of Dental Practice, Characteristics of Dentists in Private Practice and their Patients". In: (2012).

[93] *Texas A&M Geocoding Services*. `http://geoservices.tamu.edu/Services/Geocode/`. Online Database.

[94] *2010 Rural-Urban Commuting Area (RUCA) Codes*. `http://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation.aspx`. Online Database. 2010.

[95] American Dental Association. "Oral Health Care System: A State-by-State Analysis". In: (2015).

[96] M. Gentili et al. "Small-Area Estimation of Spatial Access to Care and Its Implications for Policy". In: *Journal of Urban Health* 92.5 (2015), pp. 864–909.

[97] Department of Health and Human Services. "State Standards for Access to Care in Medicaid managed Care". In: *OEI-02-11-00320* (2014).

[98] N. Serban. "A space-time varying coefficient model: The equity of service accessibility". In: *The Annals of Applied Statistics* 5.3 (2011), pp. 2024–2051.

[99] T.C. Buchmueller, S. Orzol, and L.D. Shore-Sheppard. "The effect of Medicaid payment rates on access to dental care among children". In: *NBER Working Paper No. 19218* (2013).

[100] S.L. Decker. "Medicaid payment levels to dentists and access to dental care among children and adolescents". In: *Journal of the American Medical Association* 306.2 (2011), pp. 187–193.

[101] T. Beazoglou et al. "Impact of fee increases on dental utilization rates for children living in Connecticut and enrolled in Medicaid". In: *The Journal of the American Dental Association* 146(1) (2015), pp.52–60.

[102] National Research Council Institutes of Medicine. *Improving access to oral health care for vulnerable and underserved populations*. Report. 2011.

[103] Siegal MD and Detty AM. "Do school-based dental sealant programs reach higher risk children?" In: *J Pub Health Dent* 70.13 (2010), pp. 181–7.

[104] Georgia General Assembly. "2017-2018 Regular Session: HB 154 Dental hygienists; perform certain functions under general supervision; authorize." In: (2017).

[105] *Geographic Access to Dental Care*. http://www.ada.org/en/science-research/health-policy-institute/geographic-access-to-dental-care. Online Database.

[106] Kamyar Nasseh, Yochai Eisenberg, and Marko Vujicic. "Geographic access to dental care varies in Missouri and Wisconsin". In: *Journal of Public Health Dentistry* (2017).

[107] Z. Li, N. Serban, and J. Swann. "An optimization framework for measuring spatial access over healthcare networks". In: *BMC Health Services Research* 15.273 (2015).

[108] B.A. Dye et al. *Dental caries and sealant prevalence in children and adolescents in the United States 2011-2012*. Report. US Department of Health, Human Services, Centers for Disease Control, and Prevention, National Center for Health Statistics, 2015.

[109] A.S. Anneli et al. "Sealants for preventing dental decay in the permanent teeth". In: *The Cochrane Library* (2013).

[110] V.C. Marinho et al. "Fluoride varnishes for preventing dental caries in children and adolescents". In: *Cochrane Database Syst Rev* 7.11 (2013).

[111] B. Sen et al. "Effectiveness of preventive dental visits in reducing nonpreventive dental visits and expenditures". In: *Pediatrics* 131.6 (2013), pp. 1107–1113.

[112] F. I. Utreras. "Convergence rates for multivariate smoothing spline functions". In: *Journal of Approximation Theory* 52 (1988), pp. 1–27.